

Ecología numérica de la familia Myrtaceae en la parcela permanente de 50-ha en Barro Colorado, lago Gatún, Panamá

Rosee Aurelina Félix Méndez Estudiante, Universidad Autónoma de Santo Domingo (UASD)

La ecología de comunidades nos ayuda a comprender la manera en que se distribuyen los conjuntos de especies en la naturaleza, y los modos en que estos agrupamientos están influenciados por su ambiente y las interacciones intraespecíficas. En la isla de Barro Colorado, durante los repetidos censos que se han realizado en la parcela permanente de 50-ha, se han reportado varias especies de mirtáceas, siendo su representatividad relativamente importante en dicho enclave. Ninguna investigación ha reportado ni analizado los patrones de la comunidad de mirtáceas de la isla a la fecha. El objetivo general es conocer los rasgos básicos de la estructura y composición de la comunidad de mirtáceas en relación con factores ambientales de esta parcela. Se hicieron estudios para medir el grado de asociación, agrupamiento, diversidad y ecología espacial de esta familia con la ayuda de los paquetes de R y con los datos del octavo censo de la parcela. Las mirtáceas presentaron una riqueza de 7 especies y 4 géneros con una abundancia de 5579 individuos pertenecientes a la subtribu Myrteae. Las especies del género *Eugenia* presentaron altos grados de asociación. El agrupamiento Ward de varianza mínima sugirió la partición de 4 grupos que alcanzaron el 100% de la completitud de muestra. La diversidad de mirtáceas posee una correlación positiva con *Al*, *P*, *Ca*, *Fe* y la geomorfología de pendiente media. Las especies *Chamguava schippii* y *E. oerstediana* son las especies que aportan a la diversidad beta y están estrechamente relacionadas con los sitios (14 y 19) que aportan a la misma. El modelo de abundancia de especies mostró que el 56% de la comunidad presenta mayores valores de equidad (log normal 10% y null 46%). *C. schippii*, *E. oerstediana* y *E. nesiotica* presentaron alta correlación espacial, y otras como *Myrcia gatunensis*, *E. galalonensis* y *Psidium friedrichsthalianum* mostraron un patrón espacial aleatorio. De los hallazgos, se infiere preliminarmente que *M. gatunensis* es endémica del lago Gatún, *C. schippii* y *E. oerstediana* son basófilas y *E. nesiotica* es acidófila. Este estudio presenta nuevos conocimientos sobre las Myrteae, aunque solo se queda en estadísticos debido a la falta de información que respalde o no los hallazgos obtenidos, por lo que se necesitarán nuevos estudios de las preferencias de variables ambientales y hábitats de las mirtáceas.

Keywords: Myrtaceae, Myrteae, Ecología numérica, mirtáceas, parcela permanente de 50-ha, BCI

1 Introducción

Uno de los temas centrales en la ecología de comunidades es cómo controlar la diversidad alfa, es decir, el número de especies que pueden coexistir a escalas localizadas (e.g. 1 ha de una parcela). En las plantas esta diversidad puede alcanzar niveles extraordinarios en los bosques ecuatoriales. En el caso de los árboles de pequeños bosques tropicales, ésta puede competir con la diversidad de árboles de los bosques templados de todo un hemisferio. Otras formas de vida aumentan la diversidad alfa en los bosques tropicales, especialmente las selvas tropicales, donde los árboles representan sólo el 25% de las especies vegetales. Esta diversidad de las plantas es mayor en las selvas tropicales ecuatoriales que en cualquier otro tipo de vegetación (Wright, 2002).

La isla de Barro Colorado (BCI, por sus siglas en inglés) es un bosque tropical, la isla se formó al término del canal de Panamá en 1914, desde su creación se ha utilizado como centro de investigación debido a su gran reserva natural. Se considera monumento natural protegido por el gobierno de Panamá junto a las penínsulas Peña Blanca, Bohío, Buena Vista, Frijoles y Gigante (Smithsonian Tropical Research Institute, 2010). La parcela permanente de 50 hectáreas se encuentra en el bosque caducifolio húmedo tropical de la isla de Barro Colorado. Se estableció en 1980, desde entonces se han realizado 8 censos (aprox. 1 cada 5 años), en los cuales se toman en cuenta árboles de tallos leñosos con un diámetro a la altura del pecho (DAP) mayor a 10 mm, y como resultado en cada censo, se han identificado, censado y mapeado más de 350,000 árboles individuales (Hubbell, Condit, & Foster, 2021).

Las mirtáceas (Myrtaceae Juss) son una familia de plantas leñosas del orden Myrtales, presentes en la parcela permanente de BCI. La mayoría de las especies son árboles, también hay muchas que son arbustos o subarbustos. Algunas especies producen flores y frutos, otras raíces adventicias. Se distribuyen principalmente en zonas tropicales y templadas, con poca representación en la región africana. La familia cuenta con unos 142 géneros y más de 5.500 especies, incluyendo *Psiloxylon* y *Heteropyxis*, también pueden ser citadas por otros autores como familias monogénicas Psiloxylaceae y Heteropyxidaceae. Cabe destacar que la familia integra los árboles más altos (110-140 m) del planeta (*Eucalyptus*) y al género más numeroso (1200–1800 especies) que existe (*Syzygium*), los subarbustos rizomatosos de los géneros de la sabana (*Psidium*, *Campomanesia* y *Eugenia*), el género *Metrosideros* que contiene especies arbóreas con muchas raíces adventicias, y otros géneros son lianas trepadoras de raíces. También hay un mangle, el monotípico *Osbornia*, un pequeño árbol que carece de neumatóforos (Wilson, 2010).

Las mirtáceas de BCI pertenecen a la subtribu Myrteae. Esta subtribu posee especies que son ecológicamente importantes en muchos entornos neotropicales debido a las bayas carnosas que comen aves y mamíferos, y a las flores blancas generalistas que suministran polen y recursos a las abejas. Debido a su importancia ecológica, los investigadores muestran cada vez más interés en las Myrteae como grupo modelo para estudios evolutivos, ecológicos y de conservación en biomas neotropicales. Cerca del 70% de la diversidad de especies descritas se encuentran en solo dos géneros, *Eugenia* y *Myrcia*, y son dos de los géneros de angiospermas más grandes, con 1000 y 700 especies, respectivamente. Además, se ha demostrado sistemáticamente que estos dos géneros son hermanos de linajes pobres en especies de la tribu, lo que aumenta la disparidad de la diversidad existente entre clados estrechamente relacionados. La mayor parte de la biodiversidad existente de Myrteae está restringida al Neotrópico. En el continente americano, la mayor diversidad de especies se encuentra en las selvas tropicales, la sabana del centro, el este de Brasil, el escudo de Guayana y el Caribe; una biodiversidad menor, pero aún significativa, se encuentra en la América Central continental y en la cuenca del Amazonas de tierras bajas. La diversidad de especies es relativamente baja en las zonas subtropicales y templadas del sur de Sudamérica (Patagonia) y en los Andes de gran altitud, pero con una importante variedad de géneros endémicos.

En BCI, durante los repetidos censos que se han realizado en la parcela permanente de 50-ha desde 1981 (Hubbell et al., 2021), se han reportado varias especies de mirtáceas, siendo su representatividad relativamente importante en dicho enclave.

Mediante análisis exploratorios preliminares, se detectaron patrones singulares de distribución espacial y de diversidad. Algunas especies mostraron una distribución espacial concentrada (e.g. *Chamguava schippii*). Igualmente, varias especies mostraron preliminarmente asociación con variables de suelo y geomorfológicas. Ninguna investigación ha reportado ni analizado estos patrones a la fecha. El objetivo de este estudio es caracterizar la comunidad de mirtáceas de BCI, con ayuda de técnicas de ecología numérica (e.g. análisis de asociación, agrupamiento, diversidad y ecología espacial en relación a factores ambientales), utilizando como fuente los datos disponibles

del censo número de 8 de la parcela permanente de 50-ha.

2 Metodología

2.1 Preguntas de investigación

Medición de asociación:

-¿Se detectan especies asociadas dentro de las mirtáceas de BCI?

-¿Existe asociación entre variables ambientales/atributos? ¿Cuáles variables?

Agrupamiento:

-Los cuadros (o quadrats) de 1 hectárea, ¿se organizan en grupos discontinuos según la composición de las especies de mirtáceas?

-Si existe algún patrón, ¿es consistente con alguna variable ambiental/atributo?

-¿Hay especies indicadoras o con preferencia por determinadas condiciones ambientales/atributos?

Diversidad:

-Según los análisis de estimación de riqueza, ¿están representadas suficientemente las mirtáceas en la parcela permanente de 50-ha de BCI?

-¿Existe asociación de la diversidad alpha con variables ambientales/atributos? ¿Con cuáles?

-¿Existe contribución local o por alguna especie a la diversidad beta?

Ecología espacial:

-¿Alguna(s) especies de mirtáceas presenta(n) patrón aglomerado? ¿Cuál(es)? ¿Se asocia con alguna variable?

2.2 Ámbito geográfico

Los datos analizados fueron colectados en la parcela permanente de 50 hectáreas de la isla Barro Colorado (BCI en lo adelante). La parcela está situada en la meseta central de la isla (latitud 9° 9'N, longitud 79° 51'O), y fue establecida en 1980 por Stephen Hubbell y Robin Foster. Posee 1,000 m de largo por 500 m de ancho, se divide en 1250 cuadrantes de 20x20 m (ver figura 1). En la parcela, todos los tallos leñosos con un diámetro a la altura de pecho (DAP) mayor o igual a 1 cm se encuentran marcados, enumerados, mapeados e identificados hasta el nivel de especie. Cada 5 años, esta parcela es censada para evaluar el crecimiento, la mortalidad y para el reclutamiento de nuevas generaciones de plantas. Como resultado de estos censos se han registrado más de 300 especies de árboles, arbustos y palmas con el propósito de conocer la historia de vida de las especies, interacciones y dinámica de la comunidad (Pérez et al., 2005).

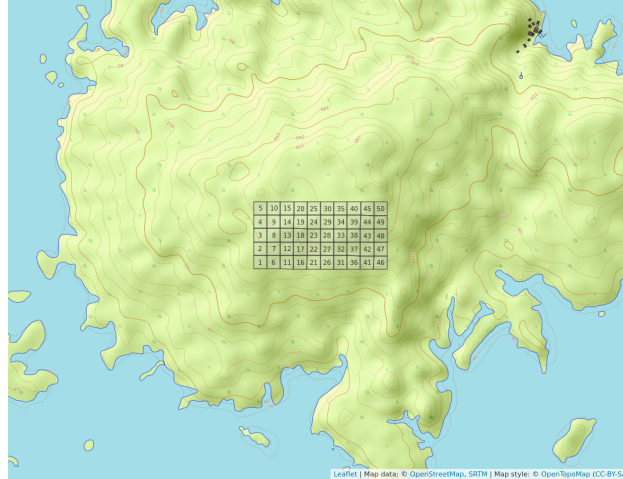


Figure 1: Parcela permanente de 50-ha de la isla Barro Colorado, lago Gatún, Panamá

2.3 Materiales y Métodos

Se ha seleccionado el censo número 8 de esta reserva natural por ser el más reciente y a esta reserva natural en particular debido a la gran cantidad disponible de datos censales que a través de la Ecología numérica nos permitirán conocer rasgos básicos de la estructura y composición de la comunidad de plantas mirtáceas en relación con factores ambientales.

Se exploraron los datos del censo número 8 disponibles en la página web del censo (Hubbell et al., 2021), organizados en dos matrices: la matriz de comunidad, la cual recopila la información referente a las especies de la parcela permanente de 50-ha, y la matriz ambiental, que contiene la información referente a las variables de suelo, geomorfológicas, litológicas y de tipo de hábitat. Los análisis, tablas, figuras y gráficos se realizaron con los scripts de análisis de José R. Martínez (Batlle, 2020) y con ayuda de los paquetes de R para análisis estadísticos y ecológicos (R Core Team, 2019), cabe destacar los paquetes *vegan* (Oksanen et al., 2019), *tidyverse* (Wickham, 2017), *sf* (Pebesma, 2018), *mapview* (Appelhans, Detsch, Reudenbach, & Woellauer, 2019) y *leaflet* (Cheng, Karambelkar, & Xie, 2018) que fueron los más utilizados.

En los análisis de medición de asociación en modo Q, se utilizaron varias distancias, como *ji-cuadrado*, normalizada, Hellinger y Jaccard. Las tres primeras son distancias euclídeas, calculadas sobre los datos transformados, apropiadas tanto para los datos cuantitativos como para los datos de presencia-ausencia; y la última, la distancia de Jaccard (D_J) se puede expresar como la proporción de especies no compartidas. La distancia de Jaccard es el complemento a 1 de la similitud de Jaccard (S_J), es decir, $D_J = 1 - S_J$, de esta manera para obtener la similitud, sólo hay que restarle el valor de distancia a 1 ($S_J = 1 - D_J$). Se puede usar para evaluar la distancia entre especies, usando como fuente la matriz de comunidad transpuesta convertida a binaria (presencia / ausencia) (Borcard, 2018).

Para el análisis de medición de asociación en modo R se utilizó el coeficiente de correlación de Pearson, el cual tiene como objetivo medir la fuerza o grado de asociación entre dos variables aleatorias cuantitativas que poseen una distribución normal bivariada conjunta. Alternativamente cuando este no cumple con los supuestos se utiliza coeficiente de correlación no paramétrico de Spearman, que se define como el coeficiente de correlación lineal entre los rangos $R_i(x)$ y $R_i(y)$ (Restrepo & González, 2007).

Se realizaron análisis de agrupamiento utilizando distintos métodos (e.g. UPGMA, *Ward*) para explorar la estructura de la comunidad en función de su composición. Para elegir entre métodos se utilizó la correlación cofenética; se consideró al agrupamiento con la correlación más alta como aquel que retiene la mayor parte de la información contenida en la matriz de disimilitud; no obstante, esto no significa necesariamente que este método sea el más adecuado para el objetivo del investigador. Luego, para escoger una cantidad óptima de clústers para cada agrupamiento se utilizó la anchura de la silueta, ésta es una medida del grado de pertenencia de un objeto a su clúster, basada en la disimilitud media entre este objeto y el clúster al que pertenece, comparada con la misma medida del clúster más próximo (Borcard, 2018).

Los métodos aglomerativos utilizados para constatar y evaluar los grupos que hacían sentido para las mirtáceas de este estudio son desarrollados a continuación:

-El método aglomerativo por enlace simple (*single*), conocido como la clasificación por vecinos más cercanos, aglomera objetos en función de sus disimilitudes más cortas entre pares: la fusión de un objeto con un grupo en un nivel de disimilitud determinado sólo requiere que un objeto de cada grupo que se aglomere esté vinculado al otro en ese nivel. En consecuencia, el dendrograma resultante de una aglomeración de enlace simple suele mostrar encadenamiento de objetos. La lista de las primeras conexiones que hacen a un objeto miembro de un clúster, o que permite la fusión de dos clústeres, se denomina cadena de conexiones primarias; esta cadena forma el árbol de expansión mínima (MST).

-El método aglomerativo por enlace completo (*complete*), conocido como la clasificación del vecino más lejano, permite que un objeto se agrupe con otro grupo sólo en la disimilitud correspondiente a la del par de objetos más distante; de esta manera con mayor motivo, todos los miembros de los dos grupos están vinculados. Un grupo admite un nuevo miembro sólo a una disimilitud correspondiente al objeto más lejano del grupo. De ello se deduce que cuánto más grande es un grupo, más difícil es aglomerarse con él. La vinculación completa resulta en muchos grupos pequeños separados que se aglomeran a grandes distancias, por lo que este método es interesante para buscar e identificar discontinuidades en los datos.

-El método de grupos de pares no ponderados con media aritmética (UPGMA, por sus siglas en inglés) es el más conocido de la familia métodos aglomerativos por enlace promedio, éstos se basan en las disimilitudes medias entre los objetos o en los centroides de los grupos. El método UPGMA permite que un objeto se una a un grupo en la media de las disimilitudes entre este objeto y todos los miembros del grupo. Cuando dos grupos se unen, lo hacen a la media de las disimilitudes entre todos los miembros de un grupo y todos los miembros del otro.

-El método de agrupación de varianza mínima de *Ward* se basa en el criterio del modelo lineal de mínimos cuadrados. Su objetivo es definir los grupos de tal manera que la suma de cuadrados dentro del grupo (es decir, el error cuadrático del ANOVA) se minimiza. La suma de errores al cuadrado dentro del grupo puede calcularse como la suma de las distancias al cuadrado entre los miembros de un grupo dividido por el número de objetos (Borcard, 2018). Este método fue seleccionado porque produce grupos con números de elementos más equilibrados, o que evita los grupos de pocos elementos.

El remuestreo *bootstrap* consiste en muestrear aleatoriamente subconjuntos de los datos y calcular la agrupación en estos subconjuntos. Luego de repetir este proceso un gran número de veces, se cuenta la proporción de los resultados de clustering replicados en los que aparece un cluster determinado. Esta proporción se denomina probabilidad *bootstrap* (BP) del cluster. Adicionalmente, se aplicó el remuestreo *bootstrap* multiescalar, utiliza muestras *bootstrap* de varios tamaños diferentes para estimar el valor p de cada conglomerado. Esta mejora produce valores p “aproximadamente insesgados” (AU) (Borcard, 2018).

Para evaluar homogeneidad de promedios de las variables ambientales entre los grupos *Ward*

y las variables ambientales, se utilizaron los análisis estadísticos ANOVA, que evalúa homogeneidad de medias, y Kruskal-Wallis, que evalúa la homogeneidad de medianas; los cuales hacen sentido para agrupamientos de 3 grupos o más (Batlle, 2020).

El análisis de especies indicadoras de los grupos *Ward* se hizo mediante el método del Valor Indicador (en lo adelante, IndVal), el cual se calcula como el producto de la especificidad de una especie para el grupo objetivo por su fidelidad al grupo. La especificidad se define por la abundancia media de la especie dentro del grupo objetivo comparada con su abundancia media en todos los grupos; la fidelidad es la proporción de sitios del grupo en el que está presente la especie. Y el análisis de especies con preferencia por hábitat se realizó mediante el coeficiente de correlación biserial puntual que mide la relación entre un vector que indica la presencia o abundancia de una especie en diferentes lugares y un vector que describe a qué grupo predefinido pertenece cada sitio (Borcard, 2018).

Para medir la diversidad alpha se utilizaron los índices de diversidad, descritos a continuación:

-La equidad de Pielou (denominada también equidad de Shannon) equivale a $J = H_1 / H_0$.

-Los tres primeros números de diversidad de Hill : $N_0 = q$ (la riqueza de especies), $N_1 = e^H$ (número de especies abundantes), y $N_1 = 1 / \lambda$ (inverso de Simpson).

-Los ratios de Hill: $E_1 = N_1 / N_0$ (versión de la equidad de Shannon) y $E_2 = N_2 / N_0$ (versión de la equidad de Simpson) (Borcard, 2018).

La equidad puede relacionarse con la forma de los modelos de abundancia de especies, estos son funciones que describen la forma de los gráficos de rango/abundancia en los que la abscisa clasifica las especies en orden de abundancia decreciente y la ordenada representa las abundancias transformadas en logaritmos. Los cuatro modelos principales son las series geométricas, logarítmicas y lognormales, y el modelo de barra rota. En este orden, la uniformidad aumenta de un modelo a otro en esta secuencia (Borcard, 2018).

Para estimar la riqueza se utilizaron los modelos de enfoque asintótico: a) paramétricos: Modelo homogéneo (estándar y MLE), este asume que todas las especies tienen las mismas probabilidades de incidencia o detección; y no paramétricos: Chao1, el cual utiliza las frecuencias de únicos y duplicados para estimar el número de especies no detectadas, Chao1-bc (forma corregida de sesgo para el estimador Chao1) e iChao1 (estimador Chao1 mejorado); ICE (Estimador de cobertura basado en la incidencia) e ICE-1 (ICE modificado para casos altamente heterogéneos); Jackknife de primer orden, el cual utiliza la frecuencia de los ejemplares únicos para estimar el número de especies no detectadas y jackknife de 2º orden, que utiliza las frecuencias de los únicos y los duplicados para estimar el número de especies no detectadas. Los cuales contienen un intervalo de confianza del 95%, para el cual se utiliza una transformación logarítmica, de modo que el límite inferior del intervalo resultante sea al menos el número de especies observadas (Chao et al., 2014).

Basado en los supuestos de Whittaker, según los cuales la diversidad beta es la variación espacial de la diversidad entre sitios dentro de un área geográfica de interés. Existen diferentes ecuaciones para medir esa variación. De la investigación sobre la diversidad beta existen dos enfoques: (1) La diversidad beta puede interpretarse como una rotación, es decir, el cambio direccional en la composición de la comunidad a lo largo de un gradiente espacial, temporal o ambiental predefinido; y (2) La diversidad beta también puede definirse como la variación de la composición de la comunidad entre unidades de muestreo, sin referencia a un gradiente explícito. Ambos conceptos entran en el ámbito de la definición de Whittaker (Borcard, 2018).

Para obtener las contribuciones locales a la diversidad beta (LCBD), se calcula primero la contribución del sitio *i* a la diversidad beta global es la suma (SS_i) de los valores centrados y al

$$SS_i = \sum_{j=1}^p s_{ij}$$

cuadrado del sitio (o fila) i en la matriz S :

y luego, la contribución relativa del sitio i a la diversidad beta, la cual es denominada LCBD,

$$LCBD_i = SS_i / SS_{Total} \quad \text{donde} \quad SS_{Total} = \sum_{i=1}^n SS_i$$

es (Borcard, 2018):

Para conocer las contribuciones de las especies a la diversidad beta (SCBD), se calcula la contribución de la especie j a la diversidad beta global es la suma (SS_j) de los valores centrados y al

$$SS_j = \sum_{i=1}^n s_{ij}$$

cuadrado de la especie (o columna) j en la matriz S :

y luego, la contribución relativa de la especie j a la diversidad beta, que eso que se conoce como

$$SCBD_j = SS_j / SS_{Total} \quad \text{donde} \quad LCBD_i = SS_i / SS_{Total}$$

SCBD, es (Borcard, 2018):

Para explorar patrones espaciales de las especies y las variables ambientales, se utilizaron tanto el correlograma, como la prueba Mantel, el índice de autocorrelación I de Moran y los mapas de indicadores locales de autocorrelación espacial (en lo adelante Mapas LISA). Serán descritos a continuación.

El correlograma es un gráfico de los valores de correlación espacial frente a las clases de distancia, que combinado con pruebas estadísticas, un correlograma permite evaluar rápidamente el tipo y el alcance de la estructura de correlación espacial de una variable (Borcard, 2018).

El Índice Global de Moran es un análisis estadístico que examina de forma integral las variaciones de la autocorrelación espacial entre valores vecinos más cercanos. Se clasifican de la siguiente manera, cuando los valores tienden a agruparse se trata de una autocorrelación espacial positiva, pero si se dispersan, es una autocorrelación negativa, y si se distribuyen de forma aleatoria, se habla de que no hay autocorrelación espacial entre los valores analizados (Bucheli, 2019). Se con-

$$I(d) = \frac{\frac{1}{W} \sum_{h=1}^n \sum_{i=1}^n w_{hi} (y_h - \bar{y}) (y_i - \bar{y})}{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}$$

struye de la siguiente manera (Borcard, 2018):

La correlación espacial en el ámbito multivariante puede evaluarse y comprobarse mediante un correlograma de Mantel. Básicamente, se calcula un estadístico de Mantel estandarizado rM (análogo al coeficiente r de Pearson) entre una matriz de disimilitud entre sitios y una matriz donde los pares de sitios que pertenecen a la misma clase de distancia reciben el valor 0 y los demás pares, el valor 1. El proceso se repite para cada clase de distancia. Cada valor de rM puede probarse mediante permutaciones. La expectativa del estadístico de Mantel para la ausencia de correlación espacial es $rM = 1/4$ (Borcard, 2018).

El Índice Local de Asociación Espacial (LISA) ayuda a identificar patrones locales de asociación espacial, descomponiendo el Índice Moran para evaluar la influencia de hábitats localizados en la estadística global, esto se visualiza utilizando los Sistemas de Información Geográfica. Este índice destaca las localizaciones con valores estadísticos significativos en los indicadores, resaltando la

presencia de puntos calientes *hot spots* o atípicos espaciales, dependiendo de los datos estadísticos analizados. El resultado es la la generación de un mapa que se denomina de agrupamiento o clúster, en este se visualiza cada unidad espacial diferenciada de sus unidades vecinas (Bucheli, 2019).

3 Resultados

La familia Myrtaceae de la parcela permanente de 50-ha de BCI presentó una abundancia de 5,579 individuos pertenecientes a 7 especies y 4 géneros de la subtribu Myrteae, de las cuales las más abundantes fueron *Eugenia galalonensis* y *E. oerstediana*, representadas con 1,975 y 1,838 individuos cada una, y las especies más raras fueron *Psidium friedrichsthalianum* y *Myrcia gatunensis*, con 58 y 56 individuos respectivamente (ver figura 1).

Table 1: Abundancia por especie de la familia Myrtaceae

Latin	n
<i>Eugenia galalonensis</i>	1975
<i>Eugenia oerstediana</i>	1838
<i>Eugenia coloradoensis</i>	609
<i>Chamguava schippii</i>	541
<i>Eugenia nesiotica</i>	502
<i>Psidium friedrichsthalianum</i>	58
<i>Myrcia gatunensis</i>	56

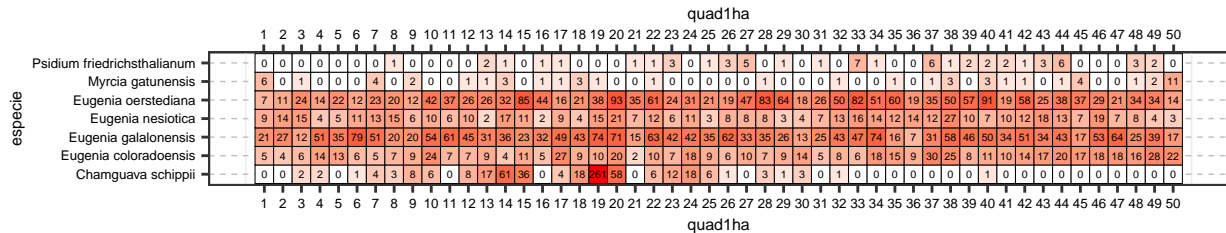


Figure 2: Abundancia de especies por quadrat

Para los análisis de medición de asociación, la distancia de *ji*-cuadrado y la distancia de Jacard resultaron pequeñas entre especies del género *Eugenia* (*E. oerstediana*, *E. galalonensis*, *E. nesiotica* y *E. coloradoensis*), lo cual sugiere un patrón de dependencia, debido a que tienen altos grados de asociación; y las especies *P. friedrichsthalianum*, *M. gatunensis* y *Chamguava schippii* presentan un posible patrón independiente, no parecen asociarse con otras (ver figura 3). La riqueza de la familia presentó asociación estadística, a través del índice de Spearman, en términos positivos con *Al*, *P* y en términos negativos con *Ca*; y la abundancia de mi familia presentó asociación estadística, a través del índice de Spearman y el índice de Pearson, en términos positivos con *Al* y elevación media, y en términos negativos con *Ca*, heterogeneidad ambiental y geomorfología de vaguada (ver figuras S1 y S2).

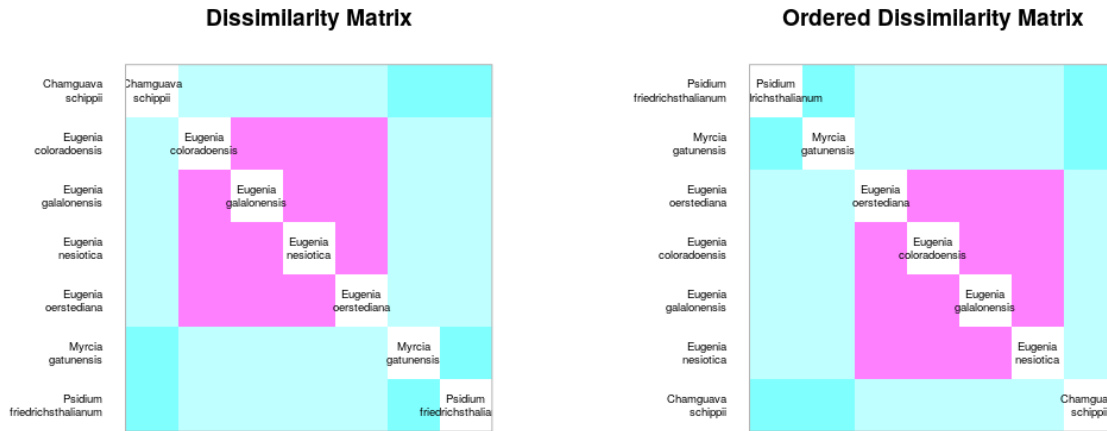


Figure 3: Matriz de disimilaridad de Jaccard

Para los análisis de agrupamiento (*cluster analysis*), se utilizó el método de agrupamiento *Ward* de varianza mínima conjuntamente con el mapa de calor (ver figuras S3 y 4), los cuales mostraron que las mirtáceas de la parcela permanente de 50-ha de BCI se distribuyen en 4 grupos, de 20, 13, 2 y 15 sitios, respectivamente (ver figura S4). Los métodos de agrupamiento aglomerativos por enlace simple, por enlace completo y por enlace promedio (grupos de pares no ponderados con media aritmética, UPGMA por sus siglas en inglés) destacaron la singularidad de este grupo formado por dos sitios (14 y 19)(ver figura S5). Además, el remuestreo de *bootstrap* multiescalar respaldó este grupo con un probabilidad de *bootstrap* (BP) de 76 % y probabilidad de valores aproximadamente insesgados (AU) de 99 %, de que sea un grupo real (ver figura S6). Las mirtáceas presentaron asociación estadística, según las pruebas ANOVA y Kruskal-Wallis para una significancia por debajo de 0.05 y el diagrama de cajas, con variables de suelo: *pH*, *Cu*, *Zn*, *Ca*, *Mg*, *N*, *Al* y *K*; y atributos del terreno: *pendiente_media*, *orientacion_media*, *elevación_media*, *geomorf_espolon/gajo_pct* (ver figuras S7 y S8).

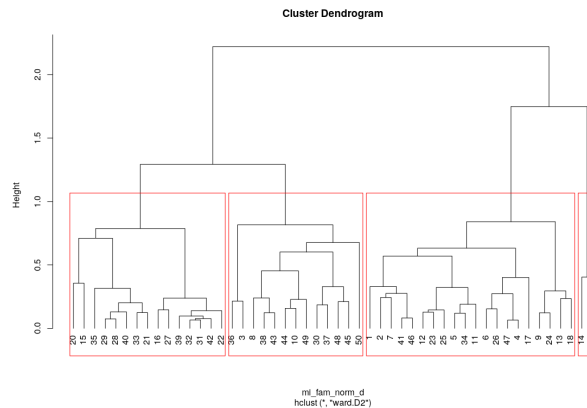


Figure 4: Agrupamiento por el método *Ward* de varianza mínima de las mirtáceas

Para este agrupamiento, las especies asociadas como indicadoras, mediante IndVal para una significancia menor de 0.05, fueron *C. schippii* para el grupo 3, *E. coloradoensis* para el conjunto de grupos 1+2 y *E. oerstediana* para el conjunto de grupos 3+4 (ver figura 5); y las especies con preferencia por hábitat, mediante el coeficiente de correlación biserial puntual para una significancia menor de 0.05, fueron *E. coloradoensis* con preferencia por el grupo 2, *C. schippii* por el grupo 3 y *E. oerstediana* por el grupo 4 (ver figura 6).

```
Multilevel pattern analysis
-----

Association function: IndVal.g
Significance level (alpha): 0.05

Total number of species: 7
Selected number of species: 3
Number of species associated to 1 group: 1
Number of species associated to 2 groups: 2
Number of species associated to 3 groups: 0

List of species associated to each combination:

Group 3 #sps. 1
          A      B  stat p.value
Chamguava schippii 0.9248 1.0000 0.962 0.002 **

Group 1+2 #sps. 1
          A      B  stat p.value
Eugenia coloradoensis 0.639 1.000 0.799 0.033 *

Group 3+4 #sps. 1
          A      B  stat p.value
Eugenia oerstediana 0.6515 1.0000 0.807 0.023 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 5: Análisis de especies indicadoras de los grupos *Ward*, mediante el método del Valor Indicador (IndVal)

```

Multilevel pattern analysis
-----

Association function: r.g
Significance level (alpha): 0.05

Total number of species: 7
Selected number of species: 3
Number of species associated to 1 group: 3
Number of species associated to 2 groups: 0
Number of species associated to 3 groups: 0

List of species associated to each combination:

Group 2 #sps. 1
              stat p.value
Eugenia coloradoensis 0.568  0.048 *

Group 3 #sps. 1
              stat p.value
Chamguava schippii  0.8    0.001 ***

Group 4 #sps. 1
              stat p.value
Eugenia oerstediana 0.74    0.01 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Figure 6: Análisis de especies con preferencia por hábitats en los grupos *Ward*, mediante el coeficiente de correlación biserial puntual

Para los análisis de diversidad alpha las mirtáceas de este ámbito geográfico presentaron, a través de la riqueza (N_0), E_2 y N_2 de Hill, una correlación positiva importante con *Al*, *P*, *Ca* y *Fe*. Además, la diversidad de la familia, a través de la equidad de Pielou (*J*) y los ratios de Hill (E_1 y E_2) y N_2 , mostró una correlación positiva notable con la presencia de la geomorfología de pendiente media (ver figura S9).

El modelo de abundancia de especies mostró que el 56% de los quadrats presenta mayores valores de equidad (log normal 10% y null 46%) (ver figura 7). También se utilizaron los modelos de estimación de riqueza (*Homogeneous model*, estándar y MLE; los Chao y los Jackknife), los cuales mostraron que la completitud de muestra se alcanzó al 100%; las estimaciones de la diversidad con muestras enrarecidas y extrapoladas mostraron que la riqueza máxima fue alcanzada (ver figura S10); y para las estimaciones de diversidad asintótica junto con estadísticas relacionadas (Riqueza de especies, diversidad de Shannon, y diversidad de Simpson) la riqueza fue estimada y observada (ver figura S11).

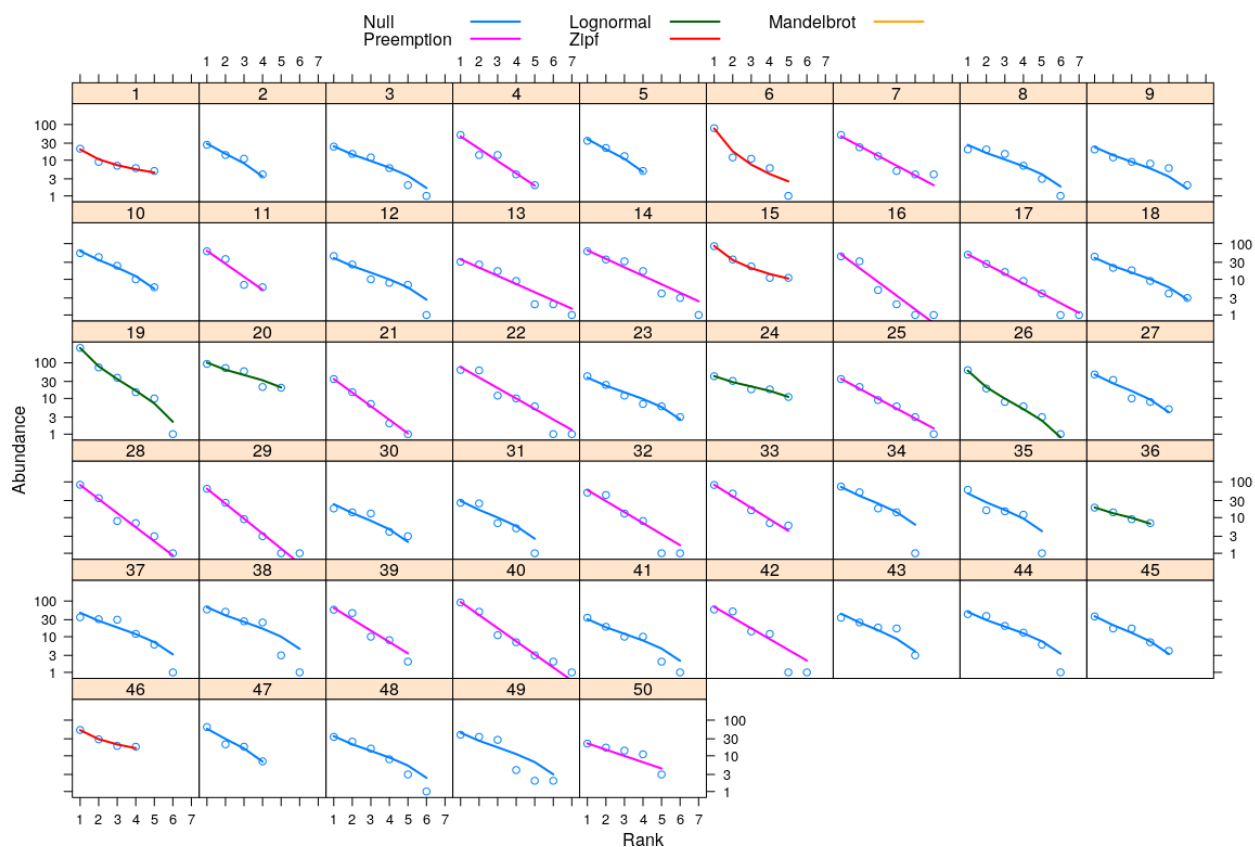


Figure 7: Modelo de abundancia de especies

La diversidad alpha en el agrupamiento *Ward*, según los modelos de estimación de riqueza (*Homogeneous model*, estándar y MLE; los Chao y los Jackknife) se alcanzó la completitud de muestra al 100% para los grupos 1, 2 y 4 (1882, 1205 y 1939 individuos, respectivamente) y al 98% para el grupo 3 (grupo con la menor abundancia, 553 individuos). Las estimaciones de la diversidad con muestras enrarecidas y extrapoladas mostraron que en los grupos *Ward* la riqueza máxima fue alcanzada (ver figuras S12, S13, S14 y S15); y para las estimaciones de diversidad asintótica junto con estadísticas relacionadas (Riqueza de especies, diversidad de Shannon, y diversidad de Simpson) la riqueza máxima fue estimada y observada (ver figura 8).

\$AsyEst: asymptotic diversity estimates along with related statistics.						
	Site	Diversity	Observed	Estimator	s.e.	LCL UCL
1	1	Species richness	7.000	7.000	0.002	7.000 7.003
2	1	Shannon diversity	4.032	4.038	0.084	4.032 4.203
3	1	Simpson diversity	3.142	3.145	0.079	3.142 3.300
4	2	Species richness	7.000	7.000	0.001	7.000 7.002
5	2	Shannon diversity	4.521	4.532	0.087	4.521 4.703
6	2	Simpson diversity	3.946	3.956	0.078	3.946 4.108
7	3	Species richness	7.000	7.000	0.500	7.000 8.446
8	3	Shannon diversity	3.329	3.348	0.128	3.329 3.600
9	3	Simpson diversity	2.508	2.515	0.130	2.508 2.770
10	4	Species richness	7.000	7.000	0.011	7.000 7.022
11	4	Shannon diversity	3.798	3.804	0.076	3.798 3.953
12	4	Simpson diversity	2.980	2.983	0.072	2.980 3.123

Figure 8: Resultado de las estimaciones de diversidad asintótica junto con estadísticas relacionadas, para el agrupamiento *Ward*

C. schippii y *E. oerstediana* fueron las especies que hicieron contribución a la diversidad beta, éstas están bien representadas (la primera con gran dominancia) en los sitios 14 y 19 (grupo 3 *Ward*) que hicieron contribución a la diversidad beta (ver figura 9); el 14 fue uno de los cinco sitios que poseen la riqueza máxima (los demás sitios son 13, 17, 22 y 40) y el 19 fue el sitio más abundante con 399 individuos, de los cuales 261 pertenecían a *C. schippii* (ver figura 2). Además, los sitios 14 y 19 están ubicados uno al lado del otro geográficamente (ver figura S4).

```

$betadiv
$beta
    SStotal    BDtotal
4.63346768 0.09456056

$SCBD
    Chamguava schippii    Eugenia coloradoensis
    0.39129504            0.10315994
    Eugenia galalonensis    Eugenia nesiotica
    0.12132302            0.08307379
    Eugenia oerstediana    Myrcia gatunensis
    0.14756292            0.08890554
    Psidium friedrichsthalianum
    0.06467975

$LCBD
    1      2      3      4      5      6
0.034502324 0.021083593 0.016380978 0.017149450 0.010164031 0.030451557
    7      8      9      10     11     12
0.013237353 0.009794359 0.018989300 0.005199322 0.013617440 0.008757331
    13     14     15     16     17     18
0.025849901 0.064059698 0.042113692 0.016894631 0.015254314 0.026597337
    19     20     21     22     23     24
0.126128710 0.027468149 0.022053327 0.003464883 0.016789843 0.016345389
    25     26     27     28     29     30
0.009980862 0.017176982 0.013462996 0.016069545 0.018294037 0.013863527
    31     32     33     34     35     36
0.008251546 0.004859232 0.019761132 0.007171961 0.023576947 0.031075930
    37     38     39     40     41     42
0.018137036 0.008327898 0.010559381 0.012717132 0.009648519 0.006194592
    43     44     45     46     47     48
0.013694652 0.010827300 0.020328969 0.011361467 0.017637132 0.011136593
    49     50
0.015408352 0.048129370

```

Figure 9: Resultado de los análisis del SCBD y LCBD

En los análisis de ecología espacial, la autocorrelación espacial mediante la prueba de Mantel mostró que hay correlación espacial inducida por alguna variable en términos positivos para el primer orden y en términos negativos para el tercer y sexto orden (hasta 500 metros) (ver figura 10). Según la prueba I de Moran global aplicada a abundancia de especies transformadas sin tendencia, *C. schippii*, *E. oerstediana* y *E. nesiotica* presentaron alta correlación espacial, y otras como *M. gatunensis*, *E. galalonensis* y *P. friedrichsthalianum* mostraron un patrón espacial aleatorio (ver figuras S16 y S17). Además, comparando con los correlogramas del I de Moran, se observaron patrones espaciales muy parecidos para *C. schippii*, *B* y *Ca*; para *E. oerstediana* y *N*; y para *E. nesiotica* y *Mn* (ver figuras 11 y 12). Esto coincide con la prueba del I de Moran local (aplicado a variables ambientales) y los mapas de *clusters* LISA (aplicado a abundancias de especies transformadas sin tendencia), *C. schippii* presentó correlación espacial con los valores de abundancia altos de *Al*, y bajos de *B*, *Zn*, *Ca* y *N*; *E. oerstediana* mostró una correlación espacial con los valores de abundancia

bajos de N , $N. min.$ y pH ; y $E. nesiotica$ con valores de abundancia bajos de Mn , y altos de N , y pH (ver figuras S18, S19 y S20).

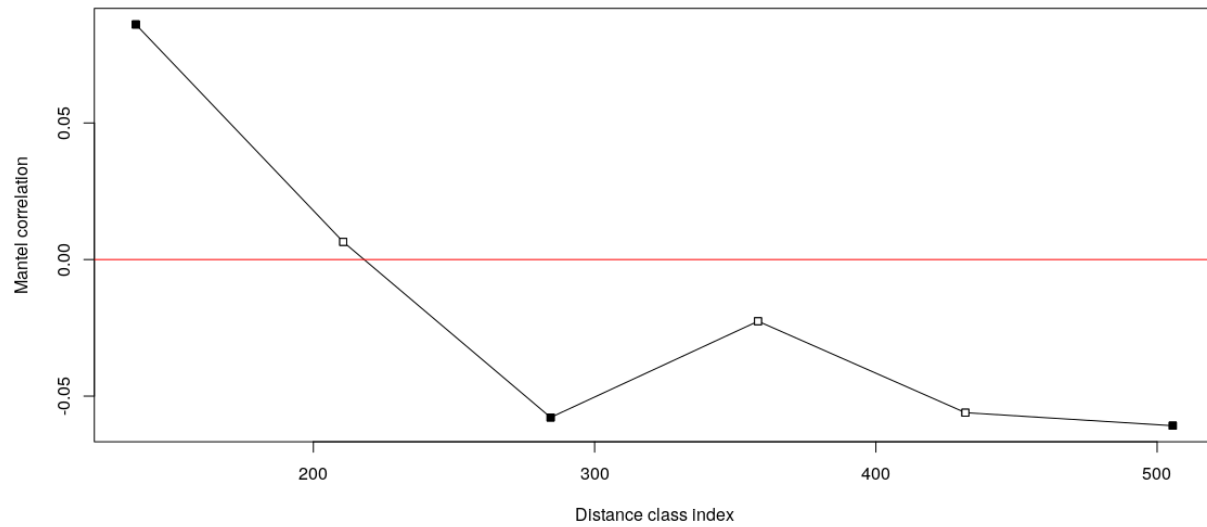


Figure 10: Autocorrelación espacial mediante prueba Mantel (matrices de distancia)

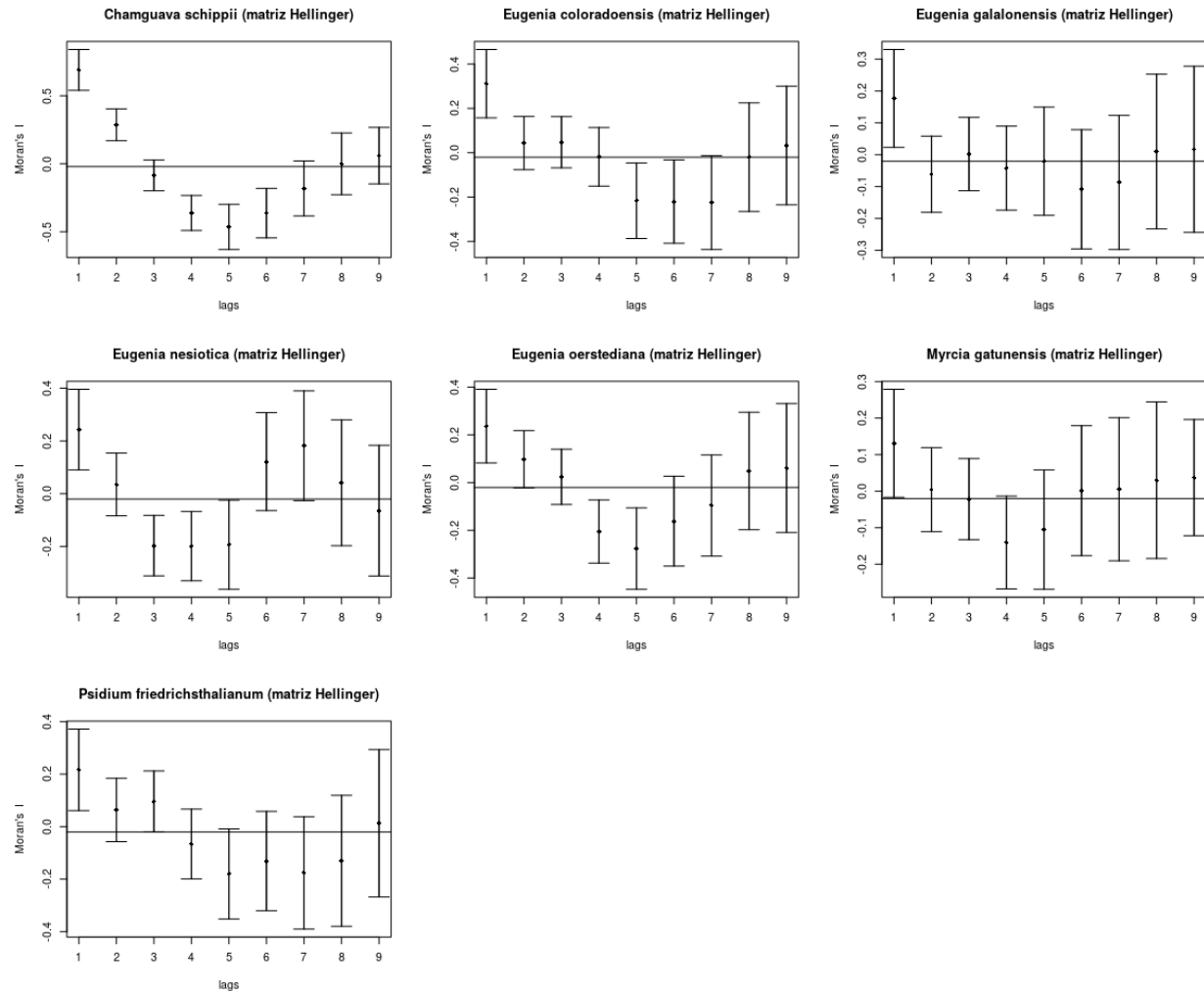


Figure 11: Correlograma, I de Moran con abundancias de especies

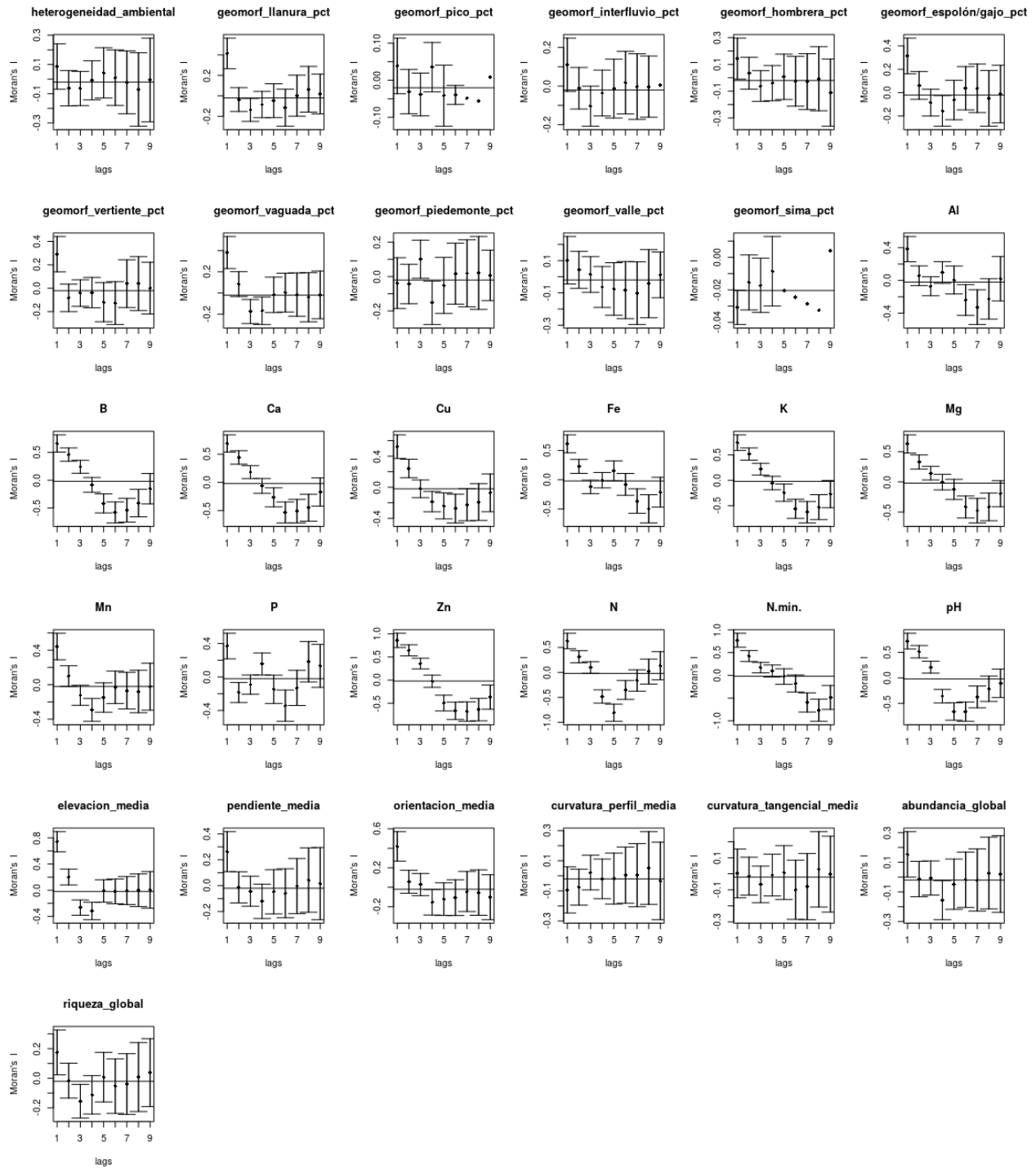


Figure 12: Correlograma, I de Moran con variables ambientales

4 Discusión

Las mirtáceas de la parcela permanente de BCI presentaron una riqueza de 7 especies y 4 géneros pertenecientes a la subtribu Myrteae. El 57% de esta riqueza, correspondiente a las especies del género *Eugenia*, presentaron altos grados de asociación entre ellas, por lo que supone un patrón de dependencia, a diferencia de las especies *P. friedrichsthalianum*, *M. gatunensis* y *C. schippii* mostraron un patrón independiente, por lo que supone que se presentan aleatoriamente en la muestra sin asociarse a las otras especies.

Las Myrteae de esta muestra, según el método *Ward*, se dividen en 4 grupos, cada uno con 20, 13, 2 y 15 sitios respectivamente. Este agrupamiento, según las pruebas ANOVA y Kruskal-Wallis y el diagrama de cajas, estuvo influenciado por variables de suelo: *pH*, *Cu*, *Zn*, *Ca*, *Mg*, *N*, *Al* y *K*; y atributos del terreno: *pendiente_media*, *orientacion_media*, *elevación_media*, *geomorf_espolon/gajo_pct*. El grupo 3 (con 2 sitios, 14 y 19) coincide en los métodos de agrupamiento (*single*, *complete*, UPGMA y *Ward*) y el remuestreo de *bootstrap* multiescalar (BP de 76% y un AU de 99%) esto infiere que es un grupo natural y real dentro de la localidad.

Según el análisis de especies indicadoras (IndVal), las especies asociadas como diagnósticas para el agrupamiento *Ward*, fueron *C. schippii* para el grupo 3, *E. coloradoensis* para el conjunto 1+2 y *E. oerstediana* para el grupo 3+4. También, *C. schippii* presentó preferencia por los hábitats del grupo 3, *E. coloradoensis* para los del grupo 2, y *E. oerstediana* para los del grupo 4. Esto puede ser debido a las altas abundancias de estas especies presentes en estos grupos.

El modelo de abundancia de especies muestra que el 56% de la comunidad presenta mayores valores de equidad (log normal 10% y null 46%). No obstante, los estimadores de riqueza mostraron que la completitud de muestra para las mirtáceas en estudio fue estimada y observada al 100%, lo mismo para los grupos *Ward*. De lo que se infiere, que este es un patrón natural de las Myrteae, y no será necesario aumentar el esfuerzo de muestreo porque no se espera encontrar más especies de esta familia en BCI.

La riqueza y abundancia de la familia presentó asociación estadística en términos positivos con *Al*, *P* y en términos negativos con *Ca*; y la abundancia presentó asociación estadística en términos positivos con *Al* y elevación media, y en términos negativos con *Ca*, heterogeneidad ambiental y geomorfología de vaguada. Además, la diversidad alpha de la comunidad presentó correlación con las variables *Al*, *P*, *Ca*, *Fe* y la geomorfología de pendiente media. De lo que se infiere, que la presencia de estas variables está relacionada estrechamente con la distribución de la familia en la parcela permanente de BCI.

Las especies que hacen contribución a la diversidad beta fueron *C. schippii* y *E. oerstediana*, y los sitios que hacen contribución fueron los sitios 14 y 19 (grupo *Ward* 3). Este grupo posee altas abundancias de las especies antes mencionadas; el sitio 19 posee la mayor abundancia de especies por *quadrat*, 399 individuos de los cuales, 261 son de *C. schippii* y 38 de *E. oerstediana*; y el sitio 14 posee 154 individuos, de los cuales 61 son de *C. schippii* y 32 de *E. oerstediana*. Estos sitios se encuentran juntos geográficamente, y presentan altos grados de *Al* y bajos grados de *N*, lo que según los análisis de ecología espacial se relaciona con altas abundancias de las especies antes mencionadas.

Los análisis de ecología espacial de mirtáceas, mediante la prueba de Mantel, mostraron estadísticamente que hay correlación espacial inducida por alguna variable en términos positivos para el primer orden. Las especies *C. schippii*, *E. oerstediana* y *E. nesiota* presentaron alta correlación espacial; y *M. gatunensis*, *E. galalonensis* y *P. friedrichsthalianum* presentaron un patrón aleatorio. Con la prueba del *I* de Moran local (aplicado a variables ambientales) y los mapas de *clusters* LISA (aplicado a abundancias de especies transformadas sin tendencia), se evidenció estadísticamente que ésta correlación para *C. schippii* está inducida por los valores de abundancia altos de *Al*, y bajos de *B*, *Zn*, *Ca* y *N*; para *E. oerstediana* por los valores de abundancia bajos de *N*,

N. min. y *pH*; y para *E. nesiotica* por los valores de abundancia bajos de *Mn*, y altos de *N*, y *pH*.

4.1 Interpretación

De los resultados obtenidos, se obtienen nuevos conocimientos preliminares de las especies de mirtáceas que se desarrollan a continuación:

-La distribución de las *Myrteae* en la parcela permanente de BCI presenta una estrecha relación con las variables *Al*, *P*, *Ca*, *Fe* y la geomorfología de pendiente media.

-*P. friedrichsthalianum* y *M. gatunensis* son especies raras que no se asocian con otras mirtáceas, presentan patrones aleatorios. Además, *M. gatunensis* fue descrita en Barro Colorado por Standl, se infiere que es preliminarmente endémica del área del lago Gatún (Kenoyer, Standley, Howe, & Dahlgren, 1929).

-*C. schippii* es una especie que no se asocia con otras mirtáceas. Su abundancia puede aumentar en presencia altos valores de *Al* y bajos *B*, *Zn*, *Ca* y *N*. Se infiere que es preliminarmente basófila.

-Las especies del género *Eugenia* presentan un patrón de dependencia. Además, se infiere que *E. oerstediana* es preliminarmente basófila y *E. nesiotica* preliminarmente acidófila.

4.2 Limitaciones y futuro

Los datos de BCI son censales, por lo que el sesgo de muestreo es una preocupación menor. Sin embargo, los datos de BCI también tienen sesgo, debido a que se utiliza un DAP de corte para decidir si un individuo es censado o no. No obstante, los datos censales carecen de una fortaleza porque no reflejan asociación con grandes unidades de hábitats y, además, revelan asociación con microhábitats muy específicos, por lo que extraer conclusiones sobre patrones de asociación con variables ambientales de manera más general, presenta sus limitaciones. Debido a esto se necesitarán más estudios de estas mirtáceas en otros hábitats menos relacionados para observar verdaderos patrones naturales y ver si coinciden con los de este estudio, que se han quedado en la estadística inferencial.

Otra de las limitaciones, es que no se encontraron estudios que respalden o no los resultados de este estudio, debido a que las preferencias de las especies de *Myrteae* de la parcela permanente de BCI han sido muy poco estudiadas. Se necesitarán más estudios sobre preferencias de variables ambientales y hábitats.

5 Agradecimientos

Primero, al profesor José R. Martínez por toda su paciencia, motivación y excelentes correcciones durante el proceso.

A mi compañero de clase, colega y amigo, Marcos A. González por leer mi trabajo, corregirme algunas faltas ortográficas, señalarme otras de sentido, motivarme y apoyarme siempre.

A mi familia, que me apoyaron, fueron pacientes y entendieron que debía pasar muchas horas despierta haciendo el trabajo y dejar de atender otras responsabilidades para alcanzar mi objetivo.

A Ruth H. Bastardo, maestra y asesora de tesis, por sus consejos y motivación siempre.

Mis más grato agradecimiento a todos, por ayudarme a concluir mi última materia de carrera.

6 Información de soporte

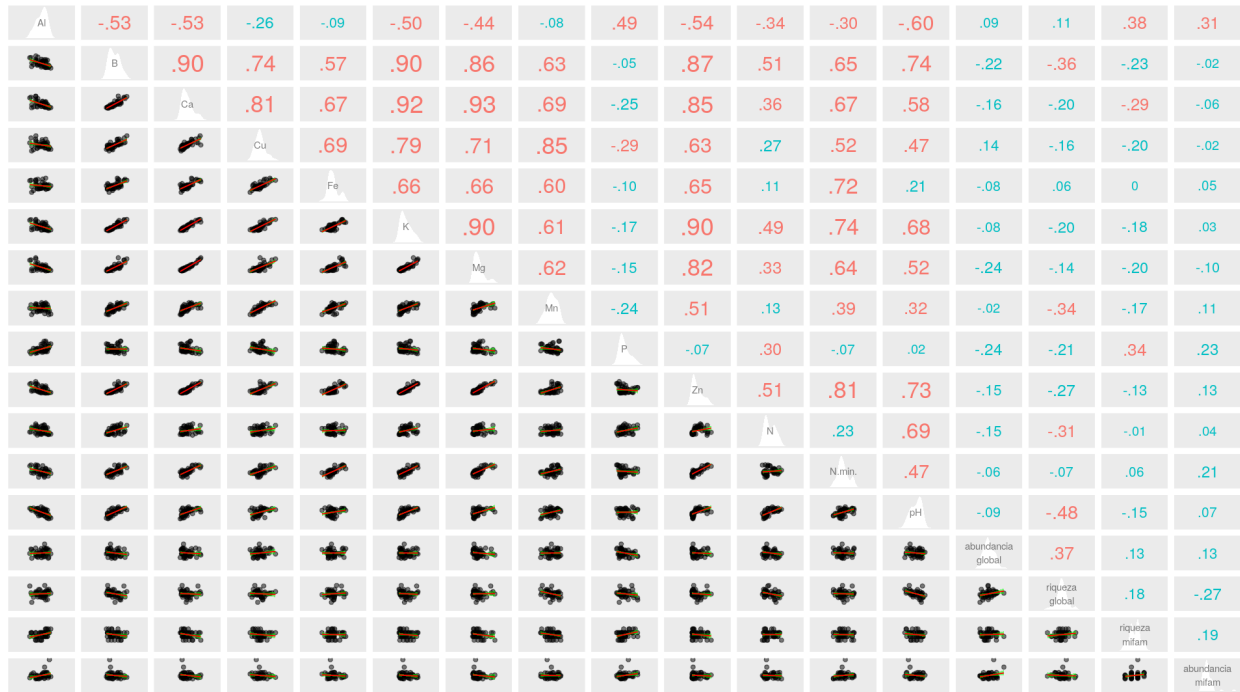


Figure S1: Matriz de correlación, índice de Spearman

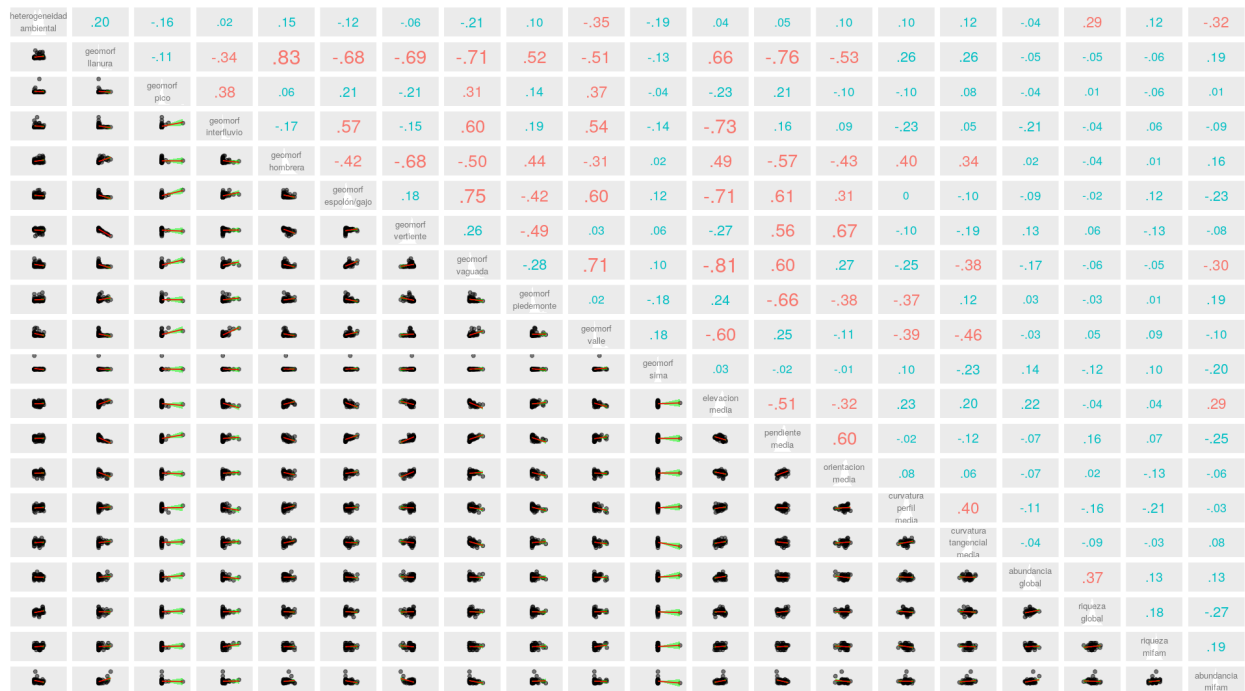


Figure S2: Matriz de correlación, índice de Pearson

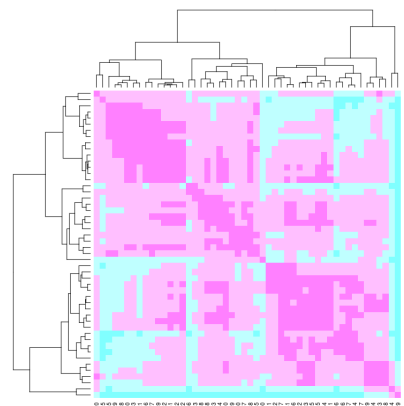


Figure S3: Mapa de calor con el dendrograma del agrupamiento *Ward*

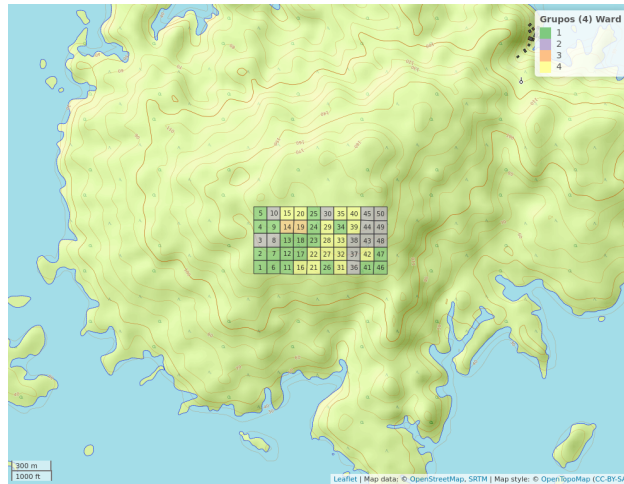


Figure S4: Agrupamiento por el método *Ward* de varianza mínima de las mirtáceas

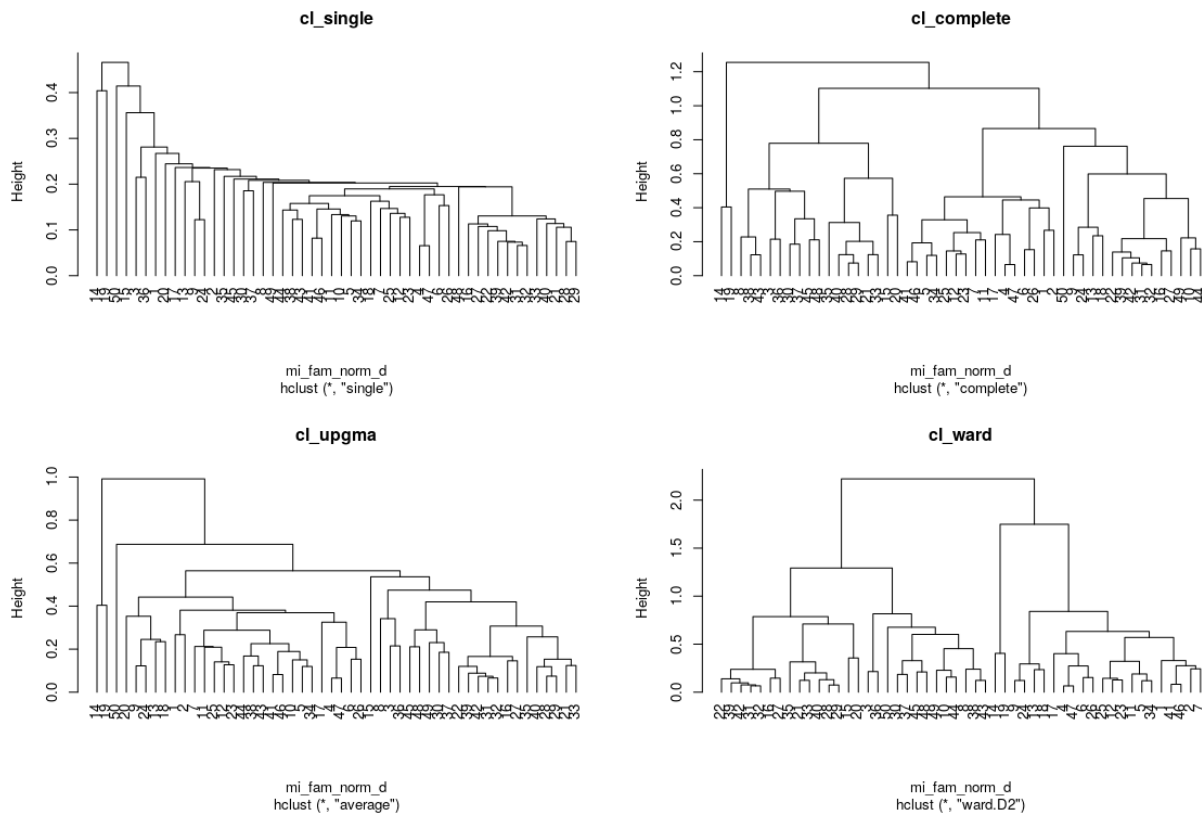


Figure S5: Métodos de agrupamiento

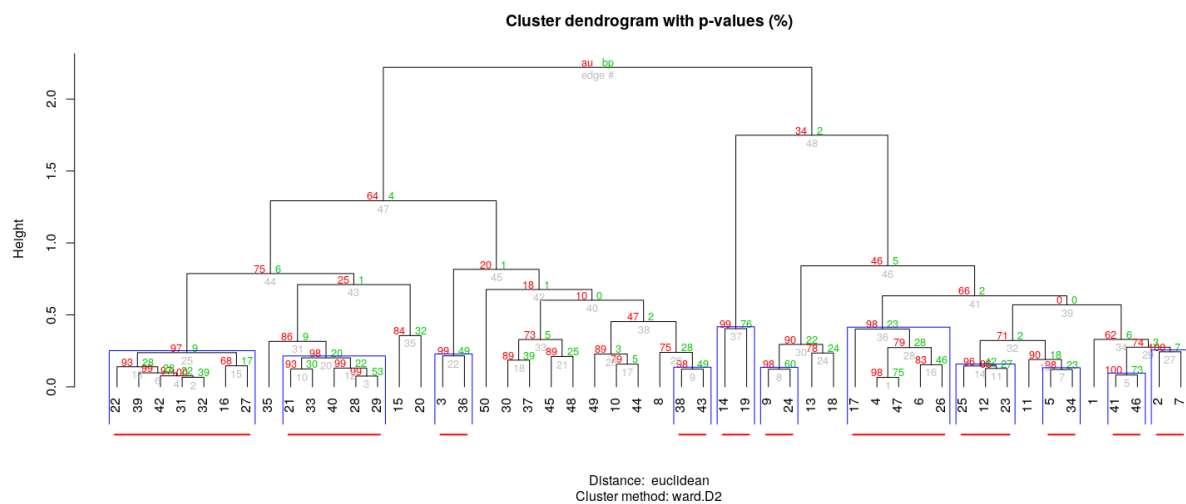


Figure S6: Dendrograma, agrupamiento *Ward* con los porcentajes del remuestreo de *bootstrap* multiescalar

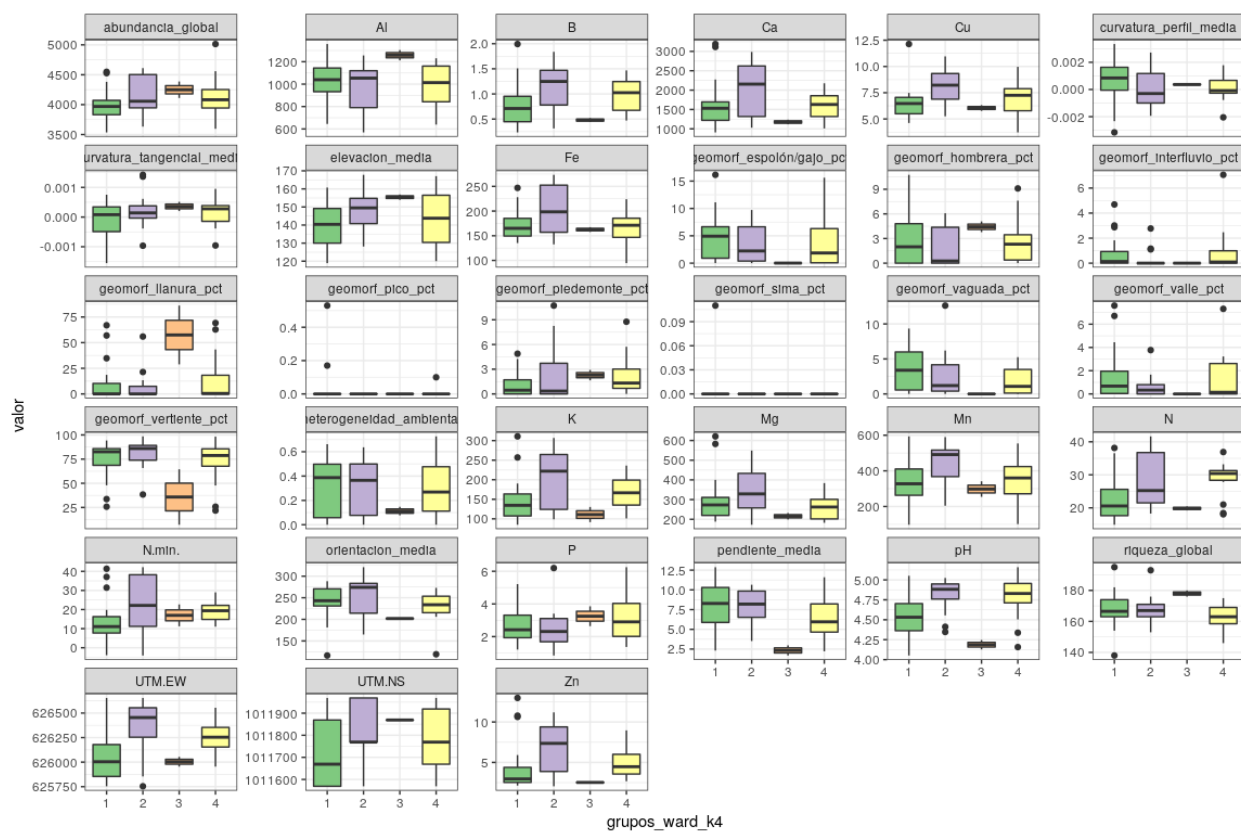


Figure S7: Diagrama de cajas de los grupos *Ward* en relación con variables ambientales y atributos

```
# A tibble: 33 x 3
  variable                p_valor_a p_valor_k
  <chr>                  <dbl>   <dbl>
1 pH                    0.000669  0.00309
2 Cu                    0.0178    0.0144
3 UTM.EW                0.0147    0.0231
4 K                     0.0865    0.0282
5 B                     0.0000313  0.0472
6 Zn                    0.00000343  0.0494
7 pendiente_media      0.00163   0.0514
8 N                     0.0000539  0.0608
9 geomorf_vaguada_pct   NaN       0.0741
10 N.min.               0.403     0.0888
11 geomorf_interfluvio_pct NaN       0.0930
12 orientacion_media    0.0000513  0.110
13 riqueza_global        0.00642   0.120
14 Ca                    0.000841  0.122
15 Mn                    0.183     0.139
16 geomorf_llanura_pct   0.493     0.142
17 geomorf_vertiente_pct 0.506     0.151
18 Mg                    0.0296   0.158
19 elevacion_media      0.000900  0.171
20 Al                    0.0168   0.179
21 UTM.NS               NaN       0.186
22 geomorf_espolón/gajo_pct 0.0000144  0.189
23 Fe                    0.0939   0.196
24 abundancia_global     0.378     0.218
25 geomorf_piedemonte_pct 0.341     0.302
26 geomorf_valle_pct     NaN       0.315
27 curvatura_perfil_media 0.560     0.420
28 geomorf_hombrera_pct  0.149     0.444
29 curvatura_tangencial_media 0.285   0.581
30 P                     0.702     0.637
31 geomorf_pico_pct      NaN       0.670
32 geomorf_sima_pct      NaN       0.682
33 heterogeneidad_ambiental 0.0142   0.753
```

Figure S8: Pruebas ANOVA y Kruskal-Wallis para el agrupamiento *Ward*

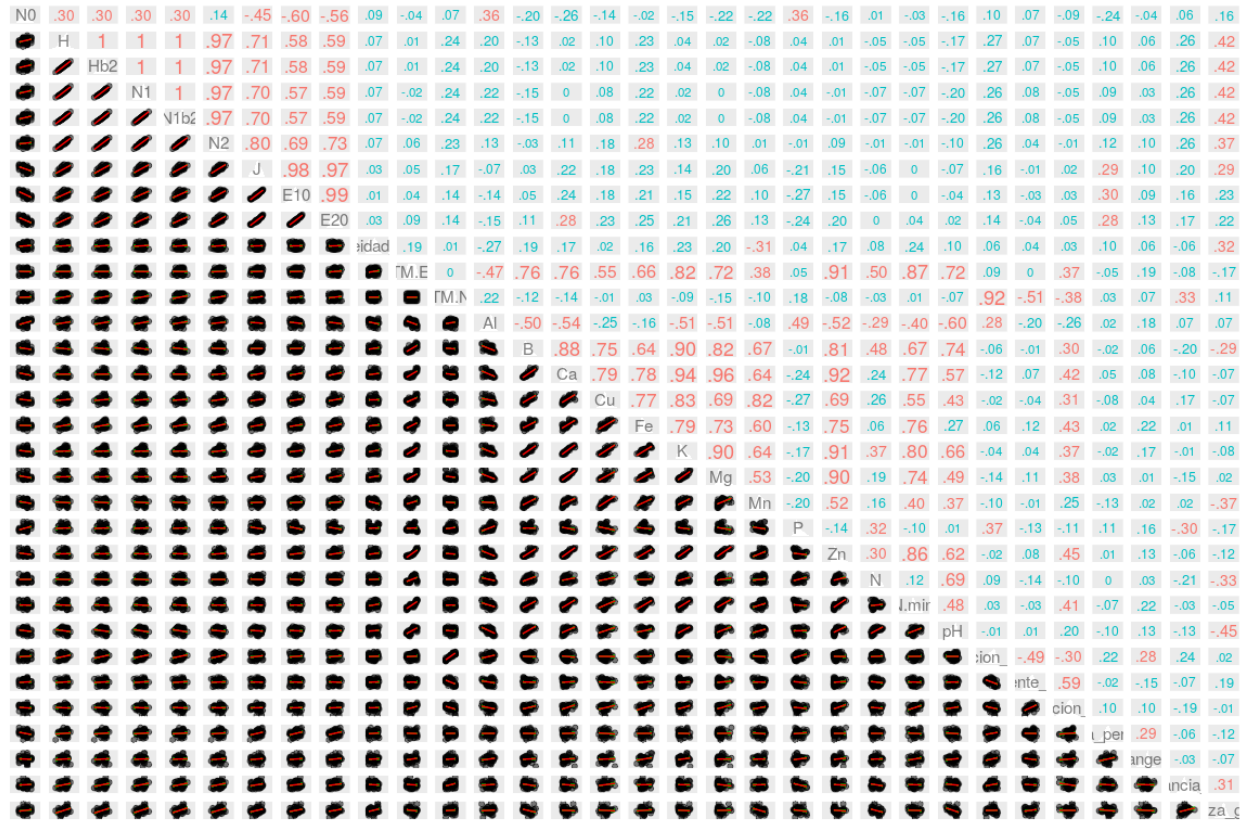


Figure S9: Matriz que relaciona los índices de diversidad alpha con las variables/atributos del suelo

(1) BASIC DATA INFORMATION:				
	Variable		Value	
Sample size	n	5579		
Number of observed species	D	7		
Coverage estimate for entire dataset	C	1		
CV for entire dataset	CV	0.918		
Cut-off point	k	60		
	Variable		Value	
Number of observed individuals for rare group	n_rare	114		
Number of observed species for rare group	D_rare	2		
Estimate of the sample coverage for rare group	C_rare	1		
Estimate of CV for rare group in ACE	CV_rare	0		
Estimate of CV1 for rare group in ACE-1	CV1_rare	0		
Number of observed individuals for abundant group	n_abun	5465		
Number of observed species for abundant group	D_abun	5		
NULL				
(2) SPECIES RICHNESS ESTIMATORS TABLE:				
	Estimate	s.e.	95%Lower	95%Upper
Homogeneous Model	7	0	7	7
Homogeneous (MLE)	7	0	7	7
Chao1 (Chao, 1984)	7	0	7	7
Chao1-bc	7	0	7	7
iChao1 (Chiu et al. 2014)	7	0	7	7
ACE (Chao & Lee, 1992)	7	0	7	7
ACE-1 (Chao & Lee, 1992)	7	0	7	7
1st order jackknife	7	0	7	7
2nd order jackknife	7	0	7	7

Figure S10: Resultado de las estimaciones de diversidad para la matriz de comunidad combinada, en la que todos los sitios forman uno

```

Compare 1 assemblages with Hill number order q = 0.
$class: iNEXT

$DataInfo: basic data information
  site    n S.obs SC f1 f2 f3 f4 f5 f6 f7 f8 f9 f10
1 site.1 5579    7 1 0 0 0 0 0 0 0 0 0 0

$iNextEst: diversity estimates with rarefied and extrapolated samples.
  m      method order qD qD.LCL qD.UCL    SC SC.LCL SC.UCL
1   1 interpolated    0 1      1      1 0.263 0.258 0.268
100 2789 interpolated    0 7      7      7 1.000 1.000 1.000
200 5579 observed    0 7      7      7 1.000 1.000 1.000
300 8354 extrapolated    0 7      7      7 1.000 1.000 1.000
400 11158 extrapolated    0 7      7      7 1.000 1.000 1.000

$AsyEst: asymptotic diversity estimates along with related statistics.
      Observed Estimator Est_s.e. 95% Lower 95% Upper
Species Richness    7.000    7.000    0.000    7.000    7.000
Shannon diversity    4.535    4.537    0.044    4.535    4.624
Simpson diversity    3.795    3.797    0.045    3.795    3.885

NOTE: Only show five estimates, call iNEXT.object$iNextEst. to show complete output.

```

Figure S11: Resultado de rarefacción y extrapolación para la matriz de comunidad combinada en la que todos los sitios forman uno

(1) BASIC DATA INFORMATION:

	Variable	Value
Sample size	n	1882
Number of observed species	D	7
Coverage estimate for entire dataset	C	1
CV for entire dataset	CV	1.107
Cut-off point	k	60

	Variable	Value
Number of observed individuals for rare group	n_rare	32
Number of observed species for rare group	D_rare	2
Estimate of the sample coverage for rare group	C_rare	1
Estimate of CV for rare group in ACE	CV_rare	0.064
Estimate of CV1 for rare group in ACE-1	CV1_rare	0.064
Number of observed individuals for abundant group	n_abun	1850
Number of observed species for abundant group	D_abun	5

NULL

(2) SPECIES RICHNESS ESTIMATORS TABLE:

	Estimate	s.e.	95%Lower	95%Upper
Homogeneous Model	7	0.002	7	7.003
Homogeneous (MLE)	7	0.002	7	7.003
Chao1 (Chao, 1984)	7	0.002	7	7.003
Chao1-bc	7	0.002	7	7.003
iChao1 (Chiu et al. 2014)	7	0.002	7	7.003
ACE (Chao & Lee, 1992)	7	0.002	7	7.003
ACE-1 (Chao & Lee, 1992)	7	0.002	7	7.003
1st order jackknife	7	0.002	7	7.003
2nd order jackknife	7	0.002	7	7.003

Figure S12: Resultado de las estimaciones de diversidad asintótica, grupo *Ward 1*

(1) BASIC DATA INFORMATION:

	Variable	Value
Sample size	n	1205
Number of observed species	D	7
Coverage estimate for entire dataset	C	1
CV for entire dataset	CV	0.877
Cut-off point	k	60

	Variable	Value
Number of observed individuals for rare group	n_rare	60
Number of observed species for rare group	D_rare	3
Estimate of the sample coverage for rare group	C_rare	1
Estimate of CV for rare group in ACE	CV_rare	0.116
Estimate of CV1 for rare group in ACE-1	CV1_rare	0.116
Number of observed individuals for abundant group	n_abun	1145
Number of observed species for abundant group	D_abun	4

NULL

(2) SPECIES RICHNESS ESTIMATORS TABLE:

	Estimate	s.e.	95%Lower	95%Upper
Homogeneous Model	7	0.001	7	7.002
Homogeneous (MLE)	7	0.001	7	7.002
Chao1 (Chao, 1984)	7	0.001	7	7.002
Chao1-bc	7	0.001	7	7.002
iChao1 (Chiu et al. 2014)	7	0.001	7	7.002
ACE (Chao & Lee, 1992)	7	0.001	7	7.002
ACE-1 (Chao & Lee, 1992)	7	0.001	7	7.002
1st order jackknife	7	0.001	7	7.002
2nd order jackknife	7	0.001	7	7.002

Figure S13: Resultado de las estimaciones de diversidad asintótica, grupo *Ward 2*

(1) BASIC DATA INFORMATION:

	Variable	Value
Sample size	n	553
Number of observed species	D	7
Coverage estimate for entire dataset	C	1
CV for entire dataset	CV	1.337
Cut-off point	k	60

	Variable	Value
Number of observed individuals for rare group	n_rare	51
Number of observed species for rare group	D_rare	4
Estimate of the sample coverage for rare group	C_rare	0.98
Estimate of CV for rare group in ACE	CV_rare	0.947
Estimate of CV1 for rare group in ACE-1	CV1_rare	1.15
Number of observed individuals for abundant group	n_abun	502
Number of observed species for abundant group	D_abun	3

NULL

(2) SPECIES RICHNESS ESTIMATORS TABLE:

	Estimate	s.e.	95%Lower	95%Upper
Homogeneous Model	7.080	0.302	7.003	9.032
Homogeneous (MLE)	7.000	0.500	7.000	8.446
Chao1 (Chao, 1984)	7.000	0.500	7.000	8.446
Chao1-bc	7.000	0.500	7.000	8.446
iChao1 (Chiu et al. 2014)	7.000	0.500	7.000	8.446
ACE (Chao & Lee, 1992)	7.996	1.869	7.090	18.066
ACE-1 (Chao & Lee, 1992)	8.430	3.010	7.112	25.303
1st order jackknife	7.998	1.412	7.128	14.792
2nd order jackknife	8.995	2.444	7.305	20.028

Figure S14: Resultado de las estimaciones de diversidad asintótica, grupo *Ward 3*

(1) BASIC DATA INFORMATION:

	Variable	Value
Sample size	n	1939
Number of observed species	D	7
Coverage estimate for entire dataset	C	1
CV for entire dataset	CV	1.161
Cut-off point	k	60

	Variable	Value
Number of observed individuals for rare group	n_rare	31
Number of observed species for rare group	D_rare	2
Estimate of the sample coverage for rare group	C_rare	1
Estimate of CV for rare group in ACE	CV_rare	0.385
Estimate of CV1 for rare group in ACE-1	CV1_rare	0.385
Number of observed individuals for abundant group	n_abun	1908
Number of observed species for abundant group	D_abun	5

NULL

(2) SPECIES RICHNESS ESTIMATORS TABLE:

	Estimate	s.e.	95%Lower	95%Upper
Homogeneous Model	7	0.011	7	7.022
Homogeneous (MLE)	7	0.011	7	7.022
Chao1 (Chao, 1984)	7	0.011	7	7.022
Chao1-bc	7	0.011	7	7.022
iChao1 (Chiu et al. 2014)	7	0.011	7	7.022
ACE (Chao & Lee, 1992)	7	0.011	7	7.022
ACE-1 (Chao & Lee, 1992)	7	0.011	7	7.022
1st order jackknife	7	0.011	7	7.022
2nd order jackknife	7	0.011	7	7.022

Figure S15: Resultado de las estimaciones de diversidad asintótica, grupo *Ward 4*

```

## '$Chamguava schippii'
##
## Monte-Carlo simulation of Moran I
##
## data: mi_fam_sin_tendencia[, x]
## weights: pesos_b
## number of simulations + 1: 10000
##
## statistic = 0.51668, observed rank = 10000, p-value = 1e-04
## alternative hypothesis: greater
##
##
## '$Eugenia coloradoensis'
##
## Monte-Carlo simulation of Moran I
##
## data: mi_fam_sin_tendencia[, x]
## weights: pesos_b
## number of simulations + 1: 10000
##
## statistic = 0.15553, observed rank = 9845, p-value = 0.0155
## alternative hypothesis: greater
##
##
## '$Eugenia galalonensis'
##
## Monte-Carlo simulation of Moran I
##
## data: mi_fam_sin_tendencia[, x]
## weights: pesos_b
## number of simulations + 1: 10000
##
## statistic = 0.0068478, observed rank = 6620, p-value = 0.338
## alternative hypothesis: greater
##
##
## '$Eugenia nesiotica'
##
## Monte-Carlo simulation of Moran I
##
## data: mi_fam_sin_tendencia[, x]
## weights: pesos_b
## number of simulations + 1: 10000
##
## statistic = 0.17369, observed rank = 9901, p-value = 0.0099
## alternative hypothesis: greater
##

```

Figure S16: I de Moran global aplicado a abundancia de especies transformadas sin tendencia, parte 1


```

## alternative hypothesis: greater
##
##
## '$Eugenia oerstediana'
##
## Monte-Carlo simulation of Moran I
##
## data: mi_fam_sin_tendencia[, x]
## weights: pesos_b
## number of simulations + 1: 10000
##
## statistic = 0.14009, observed rank = 9764, p-value = 0.0236
## alternative hypothesis: greater
##
##
## '$Myrcia gatunensis'
##
## Monte-Carlo simulation of Moran I
##
## data: mi_fam_sin_tendencia[, x]
## weights: pesos_b
## number of simulations + 1: 10000
##
## statistic = 0.099109, observed rank = 9425, p-value = 0.0575
## alternative hypothesis: greater
##
##
## '$Psidium friedrichsthalianum'
##
## Monte-Carlo simulation of Moran I
##
## data: mi_fam_sin_tendencia[, x]
## weights: pesos_b
## number of simulations + 1: 10000
##
## statistic = 0.019812, observed rank = 7187, p-value = 0.2813
## alternative hypothesis: greater

```

Figure S17: I de Moran global aplicado a abundancia de especies transformadas sin tendencia, parte 2

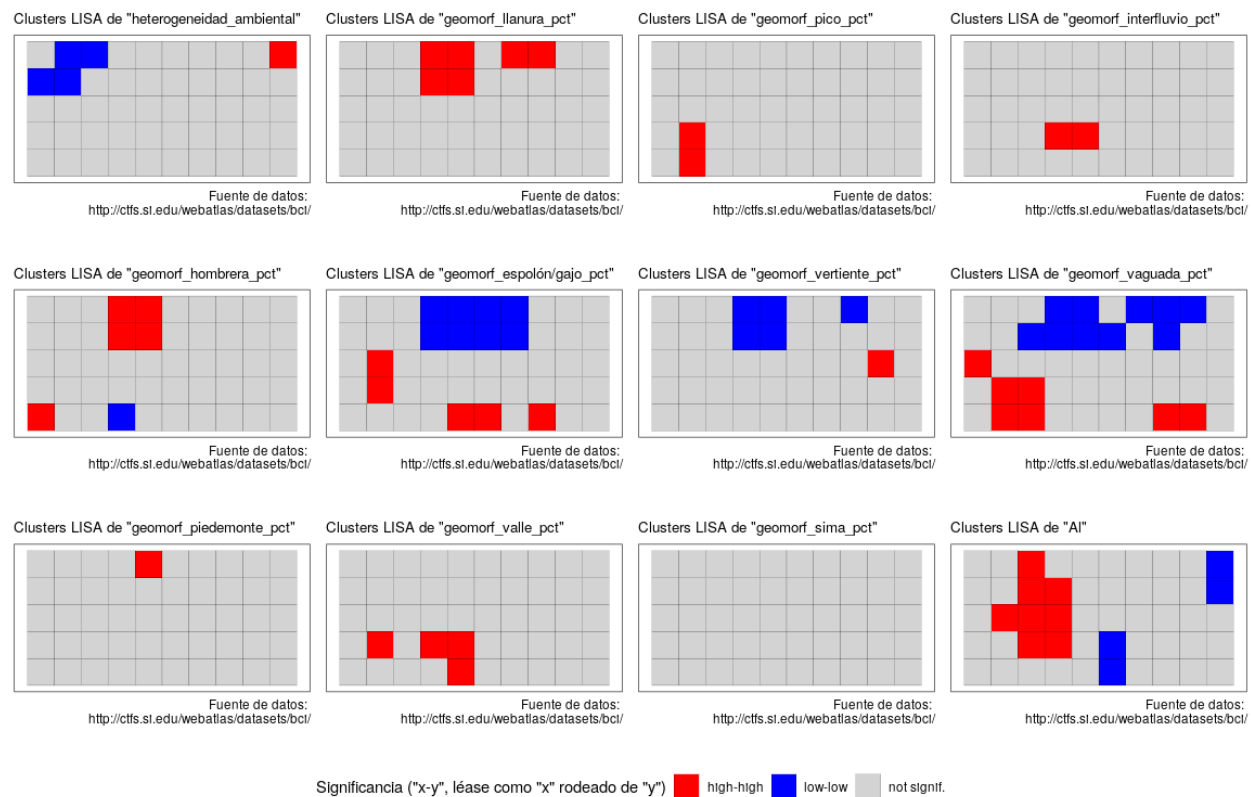


Figure S18: I de Moran local aplicado a variables ambientales, parte 1

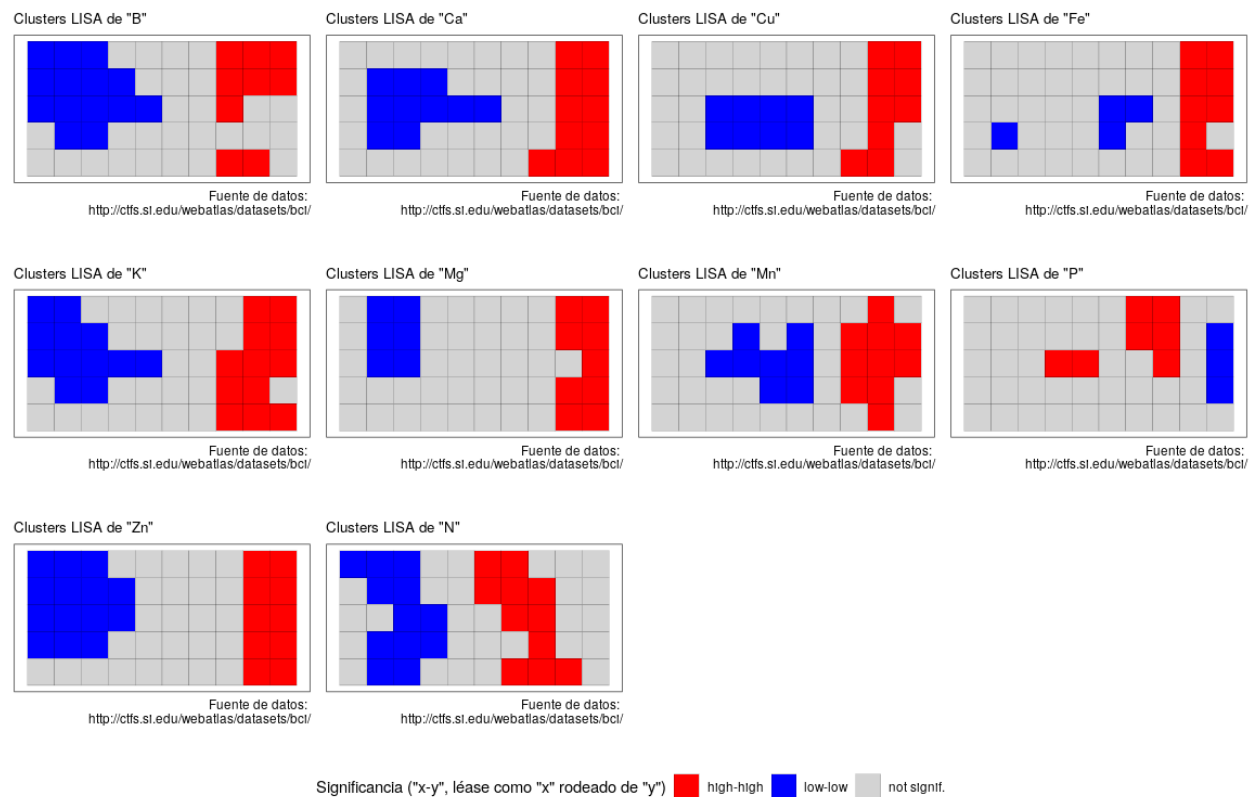


Figure S19: I de Moran local aplicado a variables ambientales, parte 2

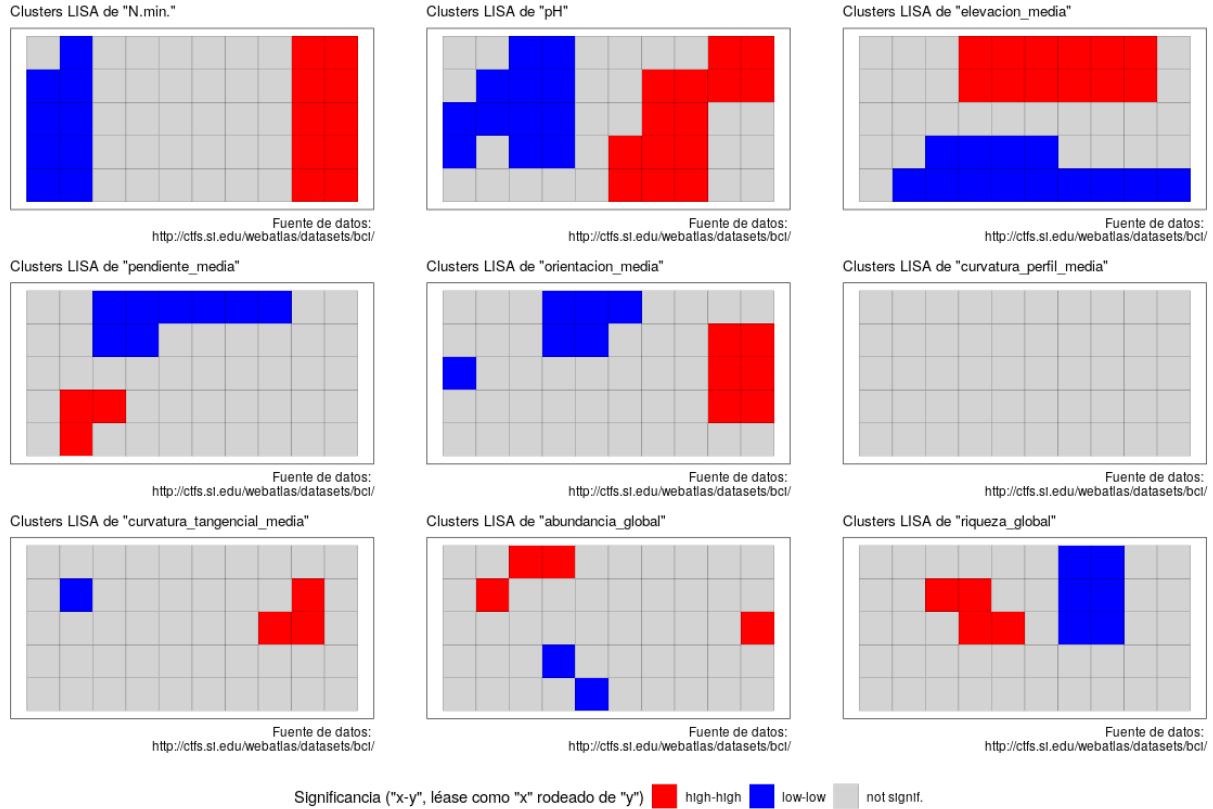


Figure S20: I de Moran local aplicado a variables ambientales, parte 3

7 Script reproducible

aed_1.R

```
#' ---
#' title: "Análisis exploratorio de datos. Riqueza y abundancia"
#' author: "JR"
#' date: "13 de octubre, 2020"
#' output: github_document
#' ---

#' ### Área de cargar paquetes
library(vegan)
library(tidyverse)
library(sf)
source('biodata/funciones.R')

#' ### Área de cargar datos
#' Censo (el objeto se carga con prefijo "censo") y matriz de comunidad (prefijo "mc")
load('biodata/Myrtaceae.Rdata')
load('biodata/matriz_ambiental.Rdata') #Matriz ambiental, se carga como "bci_env_grid"
```

```

#' ### Imprimir datos en pantalla (impresiones parciales con head)
head(censo_myrtc)
head(mc_myrtc)
bci_env_grid # No necesita imprimirse parcialmente

#' ### También podemos usar
#' Requiere que se haya cargado ya la colección tidyverse
censo_myrtc %>% tibble
mc_myrtc %>% tibble

#' ### Lista de especies
sort(colnames(mc_myrtc))

#' ### Número de sitios, tanto en matriz de comunidad como en ambiental
#' Verifica que coinciden
nrow(mc_myrtc) #En la matriz de comunidad
nrow(bci_env_grid) #En la matriz ambiental

#' ### Riqueza numérica de especies (usando matriz de comunidad) por quadrat
#' Nota: cargar paquete vegan arriba, en el área de paquetes
specnumber(mc_myrtc)
sort(specnumber(mc_myrtc)) # Ordenados ascendentemente
summary(specnumber(mc_myrtc)) # Resumen estadístico

#' ### Abundancia de especies por quadrat
sort(rowSums(mc_myrtc))
summary(rowSums(mc_myrtc)) # Resumen estadístico

#' ### Abundancia por especie
sort(colSums(mc_myrtc))
summary(colSums(mc_myrtc)) # Resumen estadístico

#' ### Riqueza numérica de toda la "comunidad"
specnumber(colSums(mc_myrtc))

#' ### Abundancia de toda la comunidad
sum(colSums(mc_myrtc))

#' ### Una tabla para el manuscrito, es necesario asignarle nombre
#' Para esto, usaré la colección "tidyverse"
abun_sp <- censo_myrtc %>%
  group_by(Latin) %>%
  count() %>%
  arrange(desc(n))
abun_sp

#' ### Un gráfico para el manuscrito

```

```
#' Gráfico de mosaicos de la abundancia por especie por cuadros
abun_sp_q <- crear_grafico_mosaico_de_mc(mc_myrtc, tam_rotulo = 6)
abun_sp_q
```

aed_2.R

```
#' ---
#' title: "Análisis exploratorio de datos. Colección tidyverse"
#' author: "JR"
#' date: "18 de octubre, 2020"
#' output: github_document
#' ---

#' # ¿Qué es tidyverse?
#'
#' Es una colección de paquetes con los que podrás importar, transformar, visualizar, modelar y
#'
#' Todos estos paquetes comparten estructuras comunes. Una de las herramientas que incorpora la
#'
#' 'objeto1 %>% funcion1()' es equivalente a 'funcion1(argumento1 = objeto1)'
#'
#' > La idea del *pipe* pertenece a la tradición de sistemas tipo Unix y, en origen, su función
#'
#' Su ventaja radica en que, si necesitaras continuar procesando los datos, no tendrás que anidar
#'
#' 'resultado <- objeto1 %>% funcion1() %>% funcion2() %>% funcion3()'
#'
#' Puedes leer lo anterior como "objeto1 pasa como primer argumento de funcion1, **luego** el r
#'
#' Para replicar esta operación sin la pipa, podrías realizarlo de, por ejemplo, dos maneras dis
#'
#' * Opción 1, anidar:
#'
#' 'resultado <- funcion3(funcion2(funcion1(objeto1)))'
#'
#' Opción 2, crear objetos intermedios:
#'
#' 'tmp1 <- funcion1(objeto1)'
#' 'tmp2 <- funcion2(tmp1)'
#' 'resultado <- funcion3(tmp2)'
#'
#' Notarás que, comparada con estas dos últimas opciones, la tubería es más limpia que estas dos
#'
#' ¡Comencemos!
#'
```

```

#' ## Paquetes
#'
```

`library(tidyverse)`
`library(sf)`

```

#'
```

*#' > 'sf' te ayudará a leer el objeto 'bci_env_grid' como un *simple feature*, el cual se encuentra en el paquete 'sf'.*

```

#'
```

#' ## Cargar datos

```

#'
```

`load('biodata/matriz_ambiental.Rdata')`
`load('biodata/Myrtaceae.Rdata')`

```

#'
```

#' ## Paquete 'dplyr'

```

#'
```

#' Te servirá para manipular datos mediante verbos. Los verbos de 'dplyr' que conocerás son (hay muchos más):

```

#'
```

#' ### Verbo 'select'

```

#'
```

#' Comúnmente, necesitas seleccionar una o varias columnas de una tabla. Para esto existe el verbo 'select'.

`bci_env_grid %>%`
 `select(id, pH)`

```

#'
```

#' > Importante: el objeto 'bci_env_grid' permanece intacto, a menos que se use dicho nombre para crear un nuevo objeto.

#' Fíjate en la clase del objeto 'bci_env_grid'. Para ello usaré la función de R 'class'. No sólo te dirá la clase, sino también la clase de cada columna.

`bci_env_grid %>%`
 `class`

```

#'
```

*#' El objeto 'bci_env_grid' es a la vez de clase 'sf' (*simple feature*) y 'data.frame', es decir, es un objeto de clase 'sf'.*

`bci_env_grid %>%`
 `select(id, pH) %>%`
 `st_drop_geometry`

```

#'
```

#' Fíjate ahora en la clase de 'bci_env_grid %>% select(id, pH) %>% st_drop_geometry', que en este caso es 'data.frame'.

`bci_env_grid %>%`
 `select(id, pH) %>%`
 `st_drop_geometry %>%`
 `class`

```

#'
```

#' > Al introducir un '<enter>' después de la pipa, el código puede continuar en la línea siguiente.

```

#'
```

#' Seleccionaré, y a la vez renombraré, dos columnas con 'select' (recuerda: no estoy modificando el objeto original).

`bci_env_grid %>%`
 `select(id_de_quadrat = id, pH_del_suelo = pH) %>%`
 `st_drop_geometry %>%`
 `head`

```

#'
```

#' Otra funcionalidad de 'select' es poder seleccionar columnas según patrones. Por ejemplo, si quiero seleccionar columnas que comiencen con 'pH'.

```

#'
```

`bci_env_grid %>%`
 `select(contains('geomorf')) %>%`
 `st_drop_geometry`

```

#’ ...y también usando expresiones regulares con ‘matches’, usando por ejemplo dos cadenas de ca
bci_env_grid %>%
  select(matches('geomorf|habit', ignore.case = F)) %>%
  st_drop_geometry
#’ ...o pidiendo todas las columnas que comienzan por mayúsculas, excepto las que comienzan por
bci_env_grid %>%
  select(matches('^[A-T,Z]', ignore.case = F)) %>%
  st_drop_geometry
#’ ### Verbo ‘filter’
#’
#’ Ahora mostraré sólo los elementos con ‘pH’ mayor que 5, usando el verbo ‘filter’
bci_env_grid %>%
  select(id, pH) %>%
  st_drop_geometry %>%
  filter(pH>5)
#’ O filtro por aquellos con ‘id’ 31 y 50:
bci_env_grid %>%
  select(id, pH) %>%
  st_drop_geometry %>%
  filter(id == c(31, 50))
#’
#’ ### Verbo ‘arrange’
#’
#’ Pruebo también con la matriz de comunidad. Por ejemplo, introduzco en la tubería la función ‘
mc_myrtc %>%
  colSums
#’ Y también obtengo la abundancia por quadrat.
mc_myrtc %>%
  rowSums
#’ Uso a continuación el verbo ‘arrange’ para mostrar los registros de la matriz ambiental orden
bci_env_grid %>%
  select(id, pH) %>%
  st_drop_geometry %>%
  arrange(pH)
#’ Ahora usaré ‘arrange’ para mostrar los registros de la matriz ambiental ordenados DESCendente
bci_env_grid %>%
  select(id, pH) %>%
  st_drop_geometry %>%
  arrange(desc(pH))
#’
#’ ### Verbo ‘mutate’
#’
#’ Usaré el verbo ‘mutate’ para crear una nueva columna. Por ejemplo, creo una columna que conte
bci_env_grid %>%
  st_drop_geometry %>%
  select(habitat, quebrada) %>%
  mutate(habitat_quebrada = paste(habitat, quebrada, sep = ', '))

```



```

#’ Ahora ‘mutate’, pero con números: creo una columna de área de cada cuadro (necesitas también
bci_env_grid %>%
  mutate(area = st_area(geometry)) %>%
  select(id, area) %>%
  st_drop_geometry %>%
  head
#’ ...y ahora más complejo: obtengo la densidad de individuos por metro cuadrado, ordenados desc
bci_env_grid %>%
  mutate(area = st_area(geometry), densidad_indiv = abundancia_global/area) %>%
  select(id, densidad_indiv) %>%
  st_drop_geometry %>%
  arrange(desc(densidad_indiv)) %>%
  head
#’
#’ ### Verbos ‘group_by’ y ‘summarise’
#’
#’ Los verbos ‘group_by’ y ‘summarise’ son útiles para producir resúmenes por grupos.
#’
#’ Agruparé la matriz ambiental por la columna ‘habitat’, dejando sólo las variables numéricas q
agrupado_por_habitat <- bci_env_grid %>%
  st_drop_geometry %>%
  group_by(habitat) %>%
  select_if(is.numeric) %>%
  select(-id, -UTM.EW, -UTM.NS)
agrupado_por_habitat
#’ Observa el encabezado: el objeto es ‘A tibble: 50 x 32’ y hay 5 grupos (‘Groups:  habitat [5
agrupado_por_habitat %>% summarise(n = n())
#’ ...y también algunos estadísticos de las columnas ‘pH’, ‘abundancia_global’ y ‘riqueza_global
agrupado_por_habitat %>%
  summarise(
    n = n(),
    media_pH = mean(pH),
    media_abundancia = mean(abundancia_global),
    media_riqueza = mean(riqueza_global)
  )
#’ ...o la media de todas las variables numéricas
agrupado_por_habitat %>%
  summarise_all(mean)
#’ ...no caben, mejor por partes
agrupado_por_habitat %>%
  summarise_all(mean) %>%
  select(1:6) %>%
  print(width=300)
agrupado_por_habitat %>%
  summarise_all(mean) %>%
  select(1,7:12) %>%
  print(width=300)

```

```

agrupado_por_habitat %>%
  summarise_all(mean) %>%
  select(1, 13:25) %>%
  print(width=300)
agrupado_por_habitat %>%
  summarise_all(mean) %>%
  select(1, 26:32) %>%
  print(width=300)
#’ ...y no sólo un estadístico, sino varios:
agrupado_por_habitat %>%
  summarise_all(
    list(
      media = mean,
      mediana = median,
      varianza = var,
      minimo = min,
      maximo = max
    )
  )
#’ Ejecuto también un ANOVA de una vía, de la ‘riqueza_global’ respecto de ‘habitat’ de tipo ‘Old*’
agrupado_por_habitat %>%
  filter(str_detect(habitat, 'Old*')) %>%
  oneway.test(formula = riqueza_global ~ habitat)
#’ El resultado sugiere que "existen 'diferencias significativas' de 'riqueza_global' entre 'habitat'"
#’
#’ Finalmente, te muestro ‘join’. Más que una función, ‘join’ es una función genérica con varios usos.
#’
#’ Obtendré una tabla con dos columnas: código identificador de quadrat de 1 ha (le llamaré ‘id’) y abundancia de cada familia.
id_abundancia_fam <- mc_myrtc %>%
  mutate(abundancia_mi_familia = rowSums(.)) %>%
  rownames_to_column(var = 'id') %>%
  mutate(id = as.numeric(id)) %>% #Numérico, garantiza compatibilidad con id de bci_env_grid
  select(id, abundancia_mi_familia)
id_abundancia_fam %>% tibble
#’ Dado que ‘id_abundancia_fam’ y ‘bci_env_grid’ comparten el campo ‘id’, a través de éste se puede unir los datos.
bci_env_grid %>%
  inner_join(y = id_abundancia_fam, by = 'id')
#’ El resultado muestra la ‘bci_env_grid’, ahora con los datos de mi familia como parte de la matriz.
#’
#’ ## ‘tidyr’
#’
#’ Te ayudará a reordenar datos, mediante transformación de su estructura, para organizarlos de una manera más útil.
#’
#’ ### Verbo ‘pivot_longer’
#’
#’ Cuando necesitas reunir varias columnas, o lo que es lo mismo, hacerlas que pivoten a lo largo de una fila, puedes usar el verbo ‘pivot_longer’.
#’

```

```

#' 
#' *Tomado de: UC Business Analytics R Programming Guide. Reshaping Your Data with tidyr. https:
#'
#' Es común realizar "reunión" columnas cuando nos interesa aplicar análisis masivos a múltiples
#'
#' Pongo un ejemplo. Por tipo de hábitat, ¿cuánto es el promedio de los porcentajes de cada uno
pivotpaso1 <- bci_env_grid %>%
  st_drop_geometry %>%
  select(matches('geomorf|habitat'))
pivotpaso1 %>% tibble
#' ...luego reunir todas las columnas de geomorfología pivotando en torno a la columna 'habitat'
pivotpaso2 <- pivotpaso1 %>%
  pivot_longer(
    cols = contains('geomorf'),
    names_to = 'variable',
    values_to = 'valor')
pivotpaso2 %>% tibble
#' ...y finalmente obtener las medias de porcentajes de geomorfología por cada grupo de habitat,
pivotpaso3 <- pivotpaso2 %>%
  group_by(habitat, variable) %>%
  summarise(media = mean(valor))
pivotpaso3 %>% arrange(habitat, desc(media)) %>% print(n=Inf)
#' 'pivot_longer' también es útil para realizar paneles de gráficos de muchas variables, como ve
#'
#' La operación contraria a 'pivot_longer' se realiza con 'pivot_wider'. Supongamos que ahora ne
#'
pivotpaso3 %>%
  ungroup() %>%
  pivot_wider(
    id_cols = habitat,
    names_from = variable,
    values_from = media)
#'
#' ## 'ggplot2'
#'
#' Te ayudará en la visualización de tus datos, utilizando gramática de gráficos.
#'
#' Un gráfico 'ggplot' utiliza capas para mostrar la información. Los objetos fuente son 'data.f
#'
#' Explicaré su uso con ejemplos, descomponiendo las partes de una sentencia 'ggplot' para fines
#'
#' Primero incluiré la función 'ggplot', para crear un espacio de coordenadas según los datos di
p0 <- ggplot(bci_env_grid)
p0
#' A continuación, definiré las variables estéticas sobre las que construiré la simbología, añad
p1 <- p0 + aes(x = abundancia_global, y = riqueza_global)
p1

```

```

# El espacio de coordenadas ya está creado, y 'ggplot2' está preparado para aceptar geometrías.
p2 <- p1 + geom_point()
p2
# Dado que en 'p1' definí las coordenadas de mapeo 'aes(x = abundancia_global, y = riqueza_glob
p3 <- p2 + geom_smooth(formula = y ~ x, method = 'lm')
p3
# En 'p3', tanto 'geom_point' como 'geom_smooth' aprovechan las coordenadas del mapeo definido
#
# Una forma alterna permite definir la capa estética dentro de la geometría con resultado idéntico
p4 <- p0 +
  geom_point(mapping = aes(x = abundancia_global, y = riqueza_global))
p4
# Esta forma tiene la ventaja de ser más corta, pero tiene la desventaja de que impide reutilizar
#
# También definiré propiedades globales del gráfico mediante temas.
p5 <- p3 + theme_bw()
p5
p6 <- p3 + theme_classic()
p6
p7 <- p3 + theme_minimal()
p7
# Con una variable categórica, se pueden estilizar los elementos del gráfico. Por ejemplo, haré
p8 <- p0 +
  geom_point(
    mapping = aes(
      x = abundancia_global,
      y = riqueza_global,
      color = habitat))
p8
#
# Ahora mostraré cómo construir el último gráfico con una sentencia de conjunto, sin reaprovechar
p9 <- ggplot(bci_env_grid) +
  geom_point(
    mapping = aes(
      x = abundancia_global,
      y = riqueza_global,
      color = habitat))
p9
# Las posibilidades de personalización de gráficos de 'ggplot2' son enormes y superan el cometido
p10 <- p0 +
  geom_boxplot(mapping = aes(x = habitat, y = abundancia_global))
p10
# Y ejemplifico también 'riqueza_global':
p11 <- p0 +
  geom_boxplot(mapping = aes(x = habitat, y = riqueza_global))
p11
# ...la cual muestra efectos más marcados que 'abundancia_global'.

```

```

#’
#’ Los dos gráficos anteriores son muy informativos, pero tienen la desventaja de que para poder
#’
#’ Necesitamos tres columnas, una con los nombres de los hábitats, otra con los nombres de las v
habitat_riqueza_abundancia <- bci_env_grid %>% st_drop_geometry %>%
  select(habitat, abundancia_global, riqueza_global) %>%
  pivot_longer(
    cols = c(abundancia_global, riqueza_global),
    names_to = 'variable',
    values_to = 'valor')
habitat_riqueza_abundancia
#’ Construiré el gráfico definiendo a ‘habitat’ en el eje ‘x’, y valor en ‘y’, mientras que usar
habitat_riqueza_abundancia %>%
  ggplot() +
  aes(x = habitat, y = valor) +
  geom_boxplot() +
  facet_wrap( ~ variable, scal = 'free_y')
#’ En resumen, usa ‘tidyverse’ para sacar el máximo provecho de tus datos. El paquete ‘dplyr’ te

```

aed_3.R

```

#’ ---
#’ title: "Análisis exploratorio de datos. Mapas de riqueza y abundancia global y de mi familia"
#’ author: "JR"
#’ date: "25 de octubre, 2020"
#’ output: github_document
#’ ---

#’ ### Cargar paquetes
library(mapview)
library(tidyverse)
library(vegan)
library(sf)
library(RColorBrewer)

#’ ### Cargar datos
load('biodata/matriz_ambiental.Rdata')
load('biodata/Myrtaceae.Rdata')

#’ ### Explorar el objeto de matriz ambiental
bci_env_grid

#’ ### Generar mapa de cuadros sin simbología
mapa_cuadros <- mapView(
  bci_env_grid,

```

```

col.regions = 'grey80',
alpha.regions = 0.3,
map.types = 'OpenTopoMap',
legend = F, zoom = 14,
zcol = 'id') %>% addStaticLabels() %>%
leaflet::setView(
  lng = -79.85136,
  lat = 9.15097,
  zoom = 15)
mapa_cuadros
mapa_cuadros %>% mapshot(file = 'mapa_cuadros.png') #Genera archivo

#' ### Paletas
azul <- colorRampPalette(brewer.pal(8, "Blues"))
rojo <- colorRampPalette(brewer.pal(8, "Reds"))

#' ### Mapa de cuadros, simbología por abundancia global
mapa_cuadros_abun_global <- mapView(
  bci_env_grid,
  layer.name = 'abundancia',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T, zoom = 14,
  col.regions = azul,
  zcol = 'abundancia_global') %>%
addStaticLabels(label = bci_env_grid$abundancia_global, textsize = "7pt") %>%
leaflet::setView(
  lng = -79.85136,
  lat = 9.15097,
  zoom = 15)
mapa_cuadros_abun_global
mapa_cuadros_abun_global %>% mapshot(file = 'mapa_cuadros_abun_global.png')

#' ### Mapa de cuadros, simbología por riqueza global
mapa_cuadros_riq_global <- mapView(
  bci_env_grid,
  layer.name = 'riqueza',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T, zoom = 14,
  col.regions = rojo,
  zcol = 'riqueza_global') %>%
addStaticLabels(label = bci_env_grid$riqueza_global, textsize = "7pt") %>%
leaflet::setView(
  lng = -79.85136,
  lat = 9.15097,
  zoom = 15)

```

```

mapa_cuadros_riq_global
mapa_cuadros_riq_global %>% mapshot(file = 'mapa_cuadros_riq_global.png')

#' ### Mapa de cuadros, simbología por abundancia de mi familia
mapa_cuadros_abun_mi_familia <- mapView(
  bci_env_grid %>% mutate(abun = rowSums(mc_myrtc)),
  layer.name = 'abundancia',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T, zoom = 14,
  col.regions = azul,
  zcol = 'abun') %>%
  addStaticLabels(label = rowSums(mc_myrtc)) %>%
  leaflet::setView(
    lng = -79.85136,
    lat = 9.15097,
    zoom = 15)
mapa_cuadros_abun_mi_familia
mapa_cuadros_abun_mi_familia %>% mapshot(file = 'mapa_cuadros_abun_mi_familia.png')

#' ### Mapa de cuadros, simbología por riqueza de mi familia
mapa_cuadros_riq_mi_familia <- mapView(
  bci_env_grid %>% mutate(riq = specnumber(mc_myrtc)),
  layer.name = 'riqueza',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T, zoom = 14,
  col.regions = rojo,
  zcol = 'riq') %>%
  addStaticLabels(label = specnumber(mc_myrtc), textsize = "12pt") %>%
  leaflet::setView(
    lng = -79.85136,
    lat = 9.15097,
    zoom = 16)
mapa_cuadros_riq_mi_familia
mapa_cuadros_riq_mi_familia %>% mapshot(file = 'mapa_cuadros_riq_mi_familia.png')

```

aed_4.R

```

#' ---
#' title: "Análisis exploratorio de datos. Mapas de variables ambientales"
#' author: "JR"
#' date: "25 de octubre, 2020"
#' output: github_document
#' ---

```

```

#' ### Cargar paquetes
library(mapview)
library(tidyverse)
library(sf)
library(RColorBrewer)

#' ### Cargar datos
load('biodata/matriz_ambiental.Rdata')

#' ### Paletas
azul <- colorRampPalette(brewer.pal(8, "Blues"))
rojo <- colorRampPalette(brewer.pal(8, "Reds"))
rojo_inv <- colorRampPalette(rev(brewer.pal(8, "Reds")))

#' ### Mapa de cuadros, simbología por pendiente
mapa_cuadros_pendiente <- mapView(
  bci_env_grid,
  layer.name = 'pendiente',
  alpha.regions = 0.4,
  map.types = 'OpenTopoMap',
  legend = T, zoom = 14,
  col.regions = rojo,
  zcol = 'pendiente_media') %>%
  addStaticLabels(label = round(bci_env_grid$pendiente_media, 1)) %>%
  leaflet::setView(
    lng = -79.85136,
    lat = 9.15097,
    zoom = 15)
mapa_cuadros_pendiente
mapa_cuadros_pendiente %>% mapshot(file = 'mapa_cuadros_pendiente.png') #Genera archivo

#' ### Mapa de cuadros, simbología por Nitrógeno
mapa_cuadros_nit <- mapView(
  bci_env_grid,
  layer.name = 'N (mg/kg)',
  alpha.regions = 0.4,
  map.types = 'OpenTopoMap',
  legend = T, zoom = 14,
  col.regions = rojo,
  zcol = 'N') %>%
  addStaticLabels(label = round(bci_env_grid$N, 1)) %>%
  leaflet::setView(
    lng = -79.85136,
    lat = 9.15097,
    zoom = 15)
mapa_cuadros_nit
mapa_cuadros_nit %>% mapshot(file = 'mapa_cuadros_nit.png')

```



```

#' ### Mapa de cuadros, simbología por pH
mapa_cuadros_ph <- mapView(
  bci_env_grid,
  layer.name = 'pH',
  alpha.regions = 0.4,
  map.types = 'OpenTopoMap',
  legend = T, zoom = 14,
  col.regions = azul,
  zcol = 'pH') %>%
  addStaticLabels(label = round(bci_env_grid$pH, 1)) %>%
  leaflet::setView(
    lng = -79.85136,
    lat = 9.15097,
    zoom = 15)
mapa_cuadros_ph
mapa_cuadros_ph %>% mapshot(file = 'mapa_cuadros_ph.png')

```

aed_5.R

```

#' ---
#' title: "Análisis exploratorio de datos. Correlaciones entre variables ambientales"
#' author: "JR"
#' date: "25 de octubre, 2020"
#' output: github_document
#' ---

knitr::opts_chunk$set(fig.width=12, fig.height=8)

#' ### Cargar paquetes
library(tidyverse)
library(sf)
library(ez)
library(psych)
library(vegan)

#' ### Cargar datos
load('biodata/matriz_ambiental.Rdata')
load('biodata/Myrtaceae.Rdata')

#' ### Una correlación simple
cor(bci_env_grid$pendiente_media, bci_env_grid$geomorf_vertiente_pct)
plot(bci_env_grid$pendiente_media, bci_env_grid$geomorf_vertiente_pct)
cor.test(bci_env_grid$pendiente_media, bci_env_grid$geomorf_vertiente_pct)

#' ### Generar objeto de columnas numéricas

```

```

#’ El objeto que generaré, denominado ‘env_num’, no tendrá las columnas ‘id’ y las de coordenada
env_num <- bci_env_grid %>%
  dplyr::select_if(is.numeric) %>%
  dplyr::select(-id, -matches('^U.*')) %>%
  st_drop_geometry %>%
  mutate(
    riqueza_mifam = specnumber(mc_myrtc),
    abundancia_mifam = rowSums(mc_myrtc)) %>%
  rename_all(gsub, pattern = '_pct$', replacement = '') %>%
  rename_all(gsub, pattern = '_| ', replacement = '\\n')
env_num %>% tibble

#’ ### Panel de correlaciones con herramientas del paquete ‘graphics’ y ‘psych’
cor(env_num)
ncol(env_num)
pairs(env_num[,sample(1:33, 15)]) # paquete graphics
env_num[,sample(1:33, 15)] %>% pairs.panels #paquete psych

#’ ### Panel de correlaciones con ‘ez’
#’
#’ ##### Todas las variables (se empasta). Comentado, sólo mostrado para fines didácticos
# p_cor_todos <- env_num %>%
#   ezCor(r_size_lims = c(4,8), label_size = 4)
# p_cor_todos

#’ ##### Sólo suelo (elementos y pH), abundancia/riqueza
p_cor_suelo_ar <- env_num %>%
  dplyr::select(matches('^([A-T,Z]|abundancia|riqueza|^pH$'), ignore.case = F)) %>%
  ezCor(r_size_lims = c(4,8), label_size = 3)
p_cor_suelo_ar

#’ ##### Sólo heterogeneidad, geomorfología, abundancia/riqueza
p_cor_geomorf_ar <- env_num %>%
  dplyr::select(-matches('^([A-T,Z]|pH$'), ignore.case = F)) %>%
  ezCor(r_size_lims = c(4,8), label_size = 3)
p_cor_geomorf_ar

#’ ##### Matriz de comunidad
p_cor_mc <- mc_myrtc %>%
  rename_all(gsub, pattern = '_| ', replacement = '\\n') %>%
  ezCor(r_size_lims = c(4,8), label_size = 3)
p_cor_mc

```

aed_6.R

```
#' ---
#' title: "Análisis exploratorio de datos. Mapas de variables ambientales por lotes"
#' author: "JR"
#' date: "3 de diciembre, 2020"
#' output: github_document
#' ---

knitr::opts_chunk$set(fig.width=12, fig.height=8)

#' ## Preámbulo
#'
#' ### Cargar paquetes
#'
library(tmap)
library(sf)
library(tidyverse)
library(RColorBrewer)
#'
#' ### Cargar datos
#'
load('biodata/matriz_ambiental.Rdata')
#'
#' ## Convertir a KML
#'
st_write(
  bci_env_grid %>% rename(Name = id),
  driver = 'KML',
  dsn = 'matriz_ambiental.kml')
st_write(
  bci_env_grid %>% rename(Name = id) %>% st_centroid(),
  driver = 'KML',
  dsn = 'matriz_ambiental_puntos.kml')
#'
#' Uní los dos archivos anteriores en un único KML nombrado como 'mapa_cuadros_1ha_para_google_e
#'
#' ## Generar mapas por lotes
#'
#' ### Variables ambientales numéricas con 'ggplot2'
#'

mapas_var_amb_num_gg <- bci_env_grid %>%
  select_if(is.numeric) %>%
  gather(variable, valor, -geometry) %>%
  group_by(variable) %>%
  mutate(
    valor = scales::rescale(valor, to = c(0, 1)),
    id = rep(1:50)) %>%
```

```

ggplot +
aes(geometry = geometry, fill = valor) +
theme(axis.text = element_blank()) +
geom_sf(lwd = 0.1, color = 'grey50', alpha = 0.8) + coord_sf() +
scale_fill_gradientn(colours = brewer.pal(11, 'BrBG')) +
geom_sf_text(aes(label = id, color = between(valor, 0.3, 0.7)), size = 1.75) +
scale_color_manual(guide = FALSE, values = c("white", "black")) +
facet_wrap(~ variable, ncol = 6) +
ggtitle('Cuadros de 1 Ha de BCI. Variables ambientales numéricas escaladas de 0 a 1')
mapas_var_amb_num_gg
#'
#' PNG
#'
png(
  filename = 'mapas_variables_ambientales_numericas.png',
  width = 1700, height = 1080, res = 150)
mapas_var_amb_num_gg
dev.off()
#'
#' ### Variables ambientales numéricas con 'tmap'
#'
mapas_var_amb_num_tmap <- bci_env_grid %>%
  select_if(is.numeric) %>%
  gather(variable, valor, -geometry) %>%
  group_by(variable) %>%
  mutate(
    valor = scales::rescale(valor, to = c(0, 1)),
    id = rep(1:50)) %>%
  tm_shape() +
  tm_polygons(col = 'valor',
    palette = brewer.pal(11, 'BrBG'),
    style = 'cont',
    legend.is.portrait = FALSE) +
  tm_facets(by = 'variable', ncol = 6, nrow = 6) +
  tm_layout(main.title="Cuadros de 1 Ha de BCI. Variables ambientales numéricas escaladas de 0 a 1",
    main.title.size = 0.7,
    legend.outside.position="bottom",
    legend.outside=TRUE,
    legend.width = 0.2,
    legend.text.size = 0.5,
    legend.stack="horizontal",
    outer.margins=0)
mapas_var_amb_num_tmap
#'
#' PNG
#'
png(

```

```

  filename = 'mapas_variables_ambientales_numericas_tmap.png',
  width = 1800, height = 1400, res = 350, pointsize = 12)
mapas_var_amb_num_tmap
dev.off()
#'
#' ### Variables ambientales nominales con 'tmap'
#'
mapas_var_amb_nom_tmap <- bci_env_grid %>%
  select_if(negate(is.numeric)) %>%
  gather(variable, valor, -geometry) %>%
  tm_shape() +
  tm_polygons(col = 'valor',
              palette = brewer.pal(8, 'Set1'),
              legend.show = T) +
  tm_facets(by = 'variable', ncol = 2, free.scales = T, free.coords = T) +
  tm_layout(main.title="Cuadros de 1 Ha de BCI. Variables ambientales nominales",
            main.title.size = 0.7,
            asp = 3.5,
            legend.text.size = 0.7)
mapas_var_amb_nom_tmap
#'
#' PNG
#'
png(
  filename = 'mapas_variables_ambientales_nominales_tmap.png',
  width = 2000, height = 1200, res = 350, pointsize = 12)
mapas_var_amb_nom_tmap
dev.off()

```

medicion_asociacion2_modoQ.R

```

#' ---
#' title: "Medición de asociación. Modo Q aplicado a mi familia asignada"
#' author: "JR"
#' date: "9 de noviembre, 2020"
#' output: github_document
#' ---

knitr::opts_chunk$set(fig.width=12, fig.height=8)

#' ## Preámbulo

#' ### Cargar paquetes
library(ez)
library(psych)

```

```

library(vegan)
library(adespatial)
library(broom)
library(tidyverse)
library(sf)
library(cluster)
library(gclus)
source('biodata/funciones.R')

#' ### Cargar datos
#'
load('biodata/matriz_ambiental.Rdata')
load('biodata/Myrtaceae.Rdata')
#'
#' ## Modo Q: matrices de disimilaridad entre objetos
#'
#' ### Modo Q para datos cuantitativos de especies (abundancia). Datos de mi familia asignada
#'
#' Aplicado a mi familia asignada de BCI, en la forma de matriz de distancia euclídea, utilizando
#'
mi_fam_d_hel <- dist.ldc(mc_myrtc, "hellinger", silent = T)
mi_fam_d_hel %>% tidy # Para evitar desbordar la consola
#'
#' Para interpretar esta matriz, es necesario representarla gráficamente. En la representación e
#'
coldiss(mi_fam_d_hel, diag = T)
#'
#' Mejorable el gráfico, quizá este es más explícito:
#'
coldissgg(mi_fam_d_hel, ordered = T, nc = 4, fsz = 0)
#'
#' Con valores de distancia sobreimpresos (se empastan un poco)
#'
coldissgg(mi_fam_d_hel, ordered = T, nc = 4, fsz = 1.5)
#'
#' Puedes guardar el gráfico usando el botón 'Export' de la pestaña 'Plots'
#'
#' Una forma alterna de guardar el gráfico es mediante funciones de R. La calidad de gráficos ex
#'
png(
  filename = 'matriz_disimilaridad_hellinger.png',
  width = 2400, height = 1200, pointsize = 32
)
coldiss(mi_fam_d_hel, diag = T)
dev.off()
#'
#' MUY IMPORTANTE. La última función, 'dev.off()', es necesaria para cerrar el dispositivo. Si n

```

```

#'
#' ### Modo Q para datos binarios (presencia/ausencia)
#'
#' Habitualmente, sólo dispones de datos de presencia/ausencia. En tales casos, existe un conjunto de métricas para calcular la distancia de Jaccard y de Sorensen "a secas", sin especificar el tipo de métrica.
#' Un error común consiste en referirse a los índices de Jaccard y de Sorensen "a secas", sin especificar el tipo de métrica.
#' Si alguna vez te enfrentas a textos donde no se especifica qué tipo de métrica se usa, te sugiero ser específico.
#' Recalco: es imprescindible declarar qué tipo de métrica estás usando. Ejemplos de redacción:
#' - Correcto: "índice de disimilitud de Jaccard", "índice de similitud de Sorensen".
#' - Incorrecto: "índice de Jaccard", "índice de Sorensen".
#'
#' A continuación, muestro cómo calcular la distancia de Jaccard ( $D_J$ ) en un objeto de tipo matrix.
#'
mi_fam_jac <- vegdist(mc_myrtc, method = 'jac', binary = T)
mi_fam_jac %>% tidy # Mostrando sólo las primeras 10 combinaciones en modo data.frame
#'
#' El argumento 'binary=T' en 'vegdist' ordena que se realice primero 'decostand(mc_apcyn_meliocarpus, method = 'bray', binary = T)'.
#' En esta matriz de disimilitud, al igual que en la anterior, un valor pequeño (rosa) significa poca distancia.
#'
coldiss(mi_fam_jac, diag = T)
png(
  filename = 'matriz_disimilitud_jaccard.png',
  width = 2400, height = 1200, pointsize = 32
)
coldiss(mi_fam_jac, diag = T)
dev.off()
#'
#' La distancia de Jaccard ( $D_J$ ) se puede expresar como "la proporción de especies que no son compartidas por los dos sitios".
#' Como la distancia de Jaccard ( $D_J$ ) es el complemento a 1 de la similitud de Jaccard ( $S_J$ ), podemos calcularla como:
#'
(1 - mi_fam_jac) %>% tidy %>% rename(similitud=distance) #Similitud
#'
#' Dado que este resultado muestra la similitud, podemos leerlo como "el sitio 1 y el 2 comparten X especies".
#' La fórmula de la similitud de Jaccard es  $S_J = a/(a+b+c)$ , donde a es el número de especies compartidas, b es el número de especies que sólo están en el sitio 1, y c es el número de especies que sólo están en el sitio 2.
#' Para obtener las variables a, b y c, usaré la función 'betadiver' del paquete 'vegan'.
#'
mi_fam_abc <- betadiver(mc_myrtc)
mi_fam_abc %>%
  map(tidy) %>%

```

```

map(slice, 1) %>%
map_df(I, .id = 'tipo') %>%
dplyr::select(tipo, n_especies=distance)
# '
# ' Puedes notar que ambos sitios comparten 11 especies (**a**), que el sitio 2 no tiene especies
# '
round(4/5*100,2) #Porcentaje de especies compartidas = similaridad
# '
# ' Con 'betadiver' también puedes calcular índices de similaridad. Por ejemplo, el Jaccard se ca
# '
betadiver(mc_myrtc, method = 'j') %>% tidy
# '
# ' No obstante, usaremos esta función en los análisis de diversidad beta más adelante.
# '
# ' Además de la distancia de Jaccard, otra distancia muy utilizada es la de Sorensen o Bray-Curt
# '
mi_fam_sor <- vegdist(mc_myrtc, method = 'bray', binary = T)
mi_fam_sor %>% tidy
coldiss(mi_fam_sor, diag = T)
# '
# ' ### Modo Q para datos cuantitativos, NO de abundancia de especies (variables ambientales)
# '
# ' En este ejemplo, usaré sólo variables de suelo, todas cuantitativas, puedes combinar con otra
# '
env_suelo_punt_z <- bci_env_grid %>%
  st_drop_geometry() %>%
  dplyr::select(matches('^([A-T,Z]|~pH$'), ignore.case = F)) %>%
  scale()
env_suelo_punt_z_d <- dist(env_suelo_punt_z)
env_suelo_punt_z_d %>% tidy
coldiss(env_suelo_punt_z_d, diag = T)
# '
# ' ### Modo Q para datos cualitativos y cuantitativos (mixtos), NO de abundancia de especies (v
# '
# ' En este ejemplo, usaré las siguientes variables mixtas (funciona igualmente para datos cualit
# '
# ' - 'heterogeneidad_ambiental'. Índice cuantitativo calculado como la diversidad de Simpson a
# '
# ' - 'habitat'. Tipo de hábitat. Asume los siguientes valores posibles: *OldHigh*, *OldLow* y *O
# '
# ' - 'quebrada'. Informa sobre si hay o no quebrada. Los valores posibles son *Yes* o *No*.
# '
env_mix <- bci_env_grid %>%
  st_drop_geometry() %>%
  dplyr::select(heterogeneidad_ambiental, habitat, quebrada)
env_mix_d <- daisy(x = env_mix, metric = 'gower')
env_mix_d %>% as.dist %>% tidy

```



```

env_mix_d %>% coldiss(diag = T)
png(
  filename = 'modoQ_datosmixtos_heterogeneidadambientalyhabitat.png',
  width = 2400, height = 1200, pointsize = 32
)
coldiss(env_mix_d, diag = T)
dev.off()
env_mix_d %>% coldissgg(ordered = T, fsz = 1)
#'

```

medición_asociación3_modor.R

```

#' ---
#' title: "Medición de asociación. Modo R aplicado a mi familia asignada"
#' author: "JR"
#' date: "3 de noviembre, 2020"
#' output: github_document
#' ---

knitr::opts_chunk$set(fig.width=12, fig.height=8)

#' ## Preámbulo

#' ### Cargar paquetes
library(vegan)
library(adespatial)
library(broom)
library(tidyverse)
library(sf)
library(gclus)
source('biodata/funciones.R')

#' ### Cargar datos
#'
load('biodata/matriz_ambiental.Rdata')
load('biodata/Myrtaceae.Rdata')
#'

#' ## Modo R: matrices de dependencia entre variables (índice de correlación)
#'

#' ### Modo R para datos cuantitativos de especies (abundancia)
#'

#' En este caso, las variables usaré los valores de abundancias de especies como variables. Es d
#'

#' Aunque se podría usar el índice de correlación como métrica de la dependencia (tal como mostr
#'

```

```

#' Por tal razón, es recomendable aplicar la transformación *Chi* a la matriz de comunidad trans
#'
#' Primero, sustituyo el caracter de espacio por un <enter> en los nombres de las especies (cara
#'
mi_fam_t <- mc_myrtc %>%
  rename_all(gsub, pattern = ' ', replacement = '\\n') %>%
  t()
mi_fam_t %>% tibble
#'
#' Segundo, transformo la matriz transpuesta usando estandarización *Chi*.
#'
mi_fam_t_chi <- decostand(mi_fam_t, "chi.square")
mi_fam_t_chi %>% tibble
#'
#' Tercero, calculo la distancia euclídea.
#'
mi_fam_t_chi_d <- dist(mi_fam_t_chi)
mi_fam_t_chi_d %>% tidy
#'
#' Finalmente, creo el "mapa de calor".
#'
coldiss(mi_fam_t_chi_d, diag = TRUE)

#'
#' En el mapa de calor **ordenado** (el de la derecha), se identifica al menos un patrón de depe
#'
#' ### Modo R para datos binarios (presencia/ausencia)
#'
#' Arriba usé la distancia de Jaccard para evaluar asociación entre sitios. Dicha métrica también
#'
mi_fam_t_jac <- vegdist(mi_fam_t, "jaccard", binary = TRUE)
mi_fam_t_jac %>% tidy
coldiss(mi_fam_t_jac, diag = TRUE)
#'
#' ### Modo R para datos cuantitativos, NO de abundancia de especies (variables ambientales)
#'
#' En modo R evaluas asociación entre descriptores, es decir, entre variables. La métrica común
#'
#' En este ejemplo, mostraré la correlación entre variables de suelo y la abundancia y riqueza g
#'
env_num <- bci_env_grid %>%
  dplyr::select_if(is.numeric) %>%
  dplyr::select(-id, -matches('^U.*')) %>%
  st_drop_geometry %>%
  mutate(
    riqueza_mifam = specnumber(mc_myrtc),
    abundancia_mifam = rowSums(mc_myrtc)) %>%

```

```

  rename_all(gsub, pattern = '_pct$', replacement = '') %>%
  rename_all(gsub, pattern = '_| ', replacement = '\\n')
env_num %>% tibble

p_cor_suelo_ar <- env_num %>%
  dplyr::select(matches('^[A-T,Z]|abundancia|riqueza|^pH$', ignore.case = F)) %>%
  ezCorM(r_size_lims = c(4,8), label_size = 3, method = 'pearson')
p_cor_suelo_ar

p_cor_suelo_ar_spearman <- env_num %>%
  dplyr::select(matches('^[A-T,Z]|abundancia|riqueza|^pH$', ignore.case = F)) %>%
  ezCorM(r_size_lims = c(4,8), label_size = 3, method = 'spearman')
p_cor_suelo_ar_spearman

png(
  filename = 'matriz_correlacion_suelo_abun_riq_spearman.png',
  width = 1920, height = 1080, res = 125
)
p_cor_suelo_ar_spearman
dev.off() #NO OLVIDAR ESTA IMPORTANTE SENTENCIA

p_cor_geomorf_ar <- env_num %>%
  dplyr::select(-matches('^[A-T,Z]|pH', ignore.case = F)) %>%
  ezCorM(r_size_lims = c(4,8), label_size = 3, method = 'pearson')
p_cor_geomorf_ar

p_cor_geomorf_ar_spearman <- env_num %>%
  dplyr::select(-matches('^[A-T,Z]|pH', ignore.case = F)) %>%
  ezCorM(r_size_lims = c(4,8), label_size = 3, method = 'spearman')
p_cor_geomorf_ar_spearman

png(
  filename = 'matriz_correlacion_geomorf_abun_riq_spearman.png',
  width = 1920, height = 1080, res = 110
)
p_cor_geomorf_ar_spearman
dev.off() #NO OLVIDAR ESTA IMPORTANTE SENTENCIA

```

aa_1.R

```

#' ---
#' title: "Análisis de agrupamiento (cluster analysis). <br> Parte 1: agrupamiento jerárquico"
#' author: "JR"
#' date: "11 de noviembre, 2020"
#' output: github_document

```

```

#' ---

knitr::opts_chunk$set(fig.width=12, fig.height=8)

#' ## Preámbulo

#' ### Cargar paquetes
library(vegan)
library(magrittr)
library(broom)
source('biodata/funciones.R')

#' ### Cargar datos
#'
load('biodata/Myrtaceae.Rdata')
mi_fam <- mc_myrtc
#'

#' ## Características de las técnicas de agrupamiento
#'
#' Las técnicas de agrupamiento se clasifican según los algoritmos que emplean y el orden de eje
#'
#' - Secuenciales o simultáneos.
#' - Por aglomeración o por división. En referencias en español encontrarás "aglomerativos" y "d
#' - Monotéticos o politéticos.
#' - Jerárquicos o no jerárquicos.
#' - Probabilísticos o no probabilísticos.
#' - Restringidos o no restringidos.
#'

#' ## Agrupamiento jerárquico
#'
#' El agrupamiento jerárquico (AJ) es una técnica de agrupamiento secuencial que consiste en la
#'
#' Dentro del AJ es frecuente usar un enfoque aglomerativo, lo cual implica aplicar algoritmos s
#'
#' El AJ aglomerativo dispone de varios algoritmos de resolución del agrupamiento por pares, que
#'
#' Normalmente, en el análisis de agrupamiento nos interesa agrupar sitios en función de sus des
#'
#' Dado que los cuadros en BCI están autocorrelacionados espacialmente, violamos el supuesto de
#'

#' ### Agrupamiento "aglomerativo" por enlace simple
#'
#' Este método utiliza, como criterio de enlace para agrupar sucesivamente pares de objetos, la
#'
#' Para aplicar este método, debes transformar la matriz de comunidad utilizando alguno de los m
#'

mi_fam_norm <- decostand(mi_fam, "normalize")

```

```

mi_fam_norm_d <- vegdist(mi_fam_norm, "euc")
mi_fam_norm_d %>% tidy
#'
#' Es importante, para garantizar consistencia a lo largo del agrupamiento, asignar los nombres
#'
attr(mi_fam_norm_d, "labels") <- rownames(mi_fam)
#'
#' Posteriormente, el agrupamiento jerárquico lo realizaré con la función 'hclust' del paquete '
#'
(cl_single <- hclust(mi_fam_norm_d, method = 'single'))
#'
#' Finalmente, el dendrograma a continuación:
plot(cl_single, labels = rownames(mi_fam), hang = -1,
     main = "Sitios de BCI según composición de especies de Myrtaceae\nEnlace simple a partir de
     xlab = 'Sitios', ylab = 'Altura')
#'
#' ### Agrupamiento "aglomerativo" por enlace completo
#'
#' En este caso, el criterio de enlace para agrupar sucesivamente pares de objetos es la menor s
#'
(cl_complete <- hclust(mi_fam_norm_d, method = 'complete'))
plot(cl_complete, labels = rownames(mi_fam), hang = -1,
     main = "Sitios de BCI según composición de especies de Myrtaceae\nEnlace completo a partir
     xlab = 'Sitios', ylab = 'Altura')
#'
#' ### Agrupamiento "aglomerativo" por enlace promedio
#'
#' En este caso, el criterio de enlace para agrupar sucesivamente pares de objetos es el promedi
#'
#' Así, dependiendo de si se media o centroide, o si se ponderan o no las distancias originales,
#'
#' Sólo crearé el dendrograma del método UPGMA.
#'
(cl_upgma <- hclust(mi_fam_norm_d, method = 'average'))
plot(cl_upgma, labels = rownames(mi_fam), hang = -1,
     main = "Sitios de BCI según composición de especies de Myrtaceae\nUPGMA a partir de matriz
     xlab = 'Sitios', ylab = 'Altura')
#'
#' ### Agrupamiento por el método de Ward de varianza mínima
#'
#' Se basa en los mismos supuestos y criterios de la regresión lineal por mínimos cuadrados, sim
#'
(cl_ward <- hclust(mi_fam_norm_d, method = 'ward.D2'))
plot(cl_ward, labels = rownames(mi_fam), hang = -1,
     main = "Sitios de BCI según composición de especies de Myrtaceae\nMétodo de Ward a partir d
     xlab = 'Sitios', ylab = 'Altura')
png(

```

```

    filename = 'sitios_de_BCI_según_composición_de_especies_de_Myrtaceae_metodo_de_Ward_a_partir_d
    width = 1920, height = 1080, res = 125
)
plot(cl_ward, labels = rownames(mi_fam), hang = -1,
     main = "Sitios de BCI según composición de especies de Myrtaceae\nMétodo de Ward a partir d
     xlab = 'Sitios', ylab = 'Altura')
dev.off()

```

aa_2.R

```

#' ---
#' title: "Análisis de agrupamiento (cluster analysis). <br> Parte 2: Interpretación y comparaci
#' author: "JR"
#' date: "11 de noviembre, 2020"
#' output: github_document
#' ---

knitr::opts_chunk$set(fig.width=12, fig.height=8)

#' ## Preámbulo
#'
#' ### Cargar paquetes
#'
library(vegan)
library(tidyverse)
library(broom)
library(cluster)
library(gclus)
library(pvclust)
library(sf)
source('biodata/funciones.R')
#'
#' ### Cargar datos
#'
load('biodata/Myrtaceae.Rdata')
mi_fam <- mc_myrtc
load('biodata/matriz_ambiental.Rdata')
mi_fam %>% tibble
bci_env_grid %>% tibble
#'
#' ### Generar matriz de distancias de cuerdas
#'
mi_fam_norm <- decostand(mi_fam, "normalize")
mi_fam_norm_d <- vegdist(mi_fam_norm, "euc")
mi_fam_norm_d %>% tidy

```

```

#'
#' ## Interpretación visual de dendrogramas
#'
#' [En el script anterior](aa_analisis_de_agrupamiento_1_jerarquico.md) realicé los dendrogramas
#'
#' La primera evaluación de los dendrogramas NO debe venir de la mano de sofisticados análisis n
#'
#' Evita concentrar tu vista en grupos extremadamente pequeños; comienza analizando el árbol des
#'
#' Anota tus impresiones, para que las compares con los resultados que posteriormente obtendrás;
#'
#' Para la exploración visual, generaré los objetos de cluster dentro de una lista:
#'
lista_cl <- list(
  cl_single = hclust(mi_fam_norm_d, method = 'single'),
  cl_complete = hclust(mi_fam_norm_d, method = 'complete'),
  cl_upgma = hclust(mi_fam_norm_d, method = 'average'),
  cl_ward = hclust(mi_fam_norm_d, method = 'ward.D2')
)
#'
#' Un plot en panel 2x2 ayuda a visualizarlos todos de manera conjunta. En tu caso, observa y co
#'
par(mfrow = c(2,2))
invisible(map(names(lista_cl), function(x) plot(lista_cl[[x]], main = x, hang = -1)))
par(mfrow = c(1,1))
#'
#' En mi caso, exceptuando el dendrograma generado por medio del enlace simple, detecto al menos
#'
#' - Un grupo pequeño, compuesto por los sitios 1, 42, 12, 21, 11, 2 y 16.
#' - Un "grupo" heterogéneo y grande, conformado por 25, 31,..., 26,..., 35,..., 34,...,32, 17,..
#'
#' Además de los grupos anteriores, detecto elementos que no forman grupos, es decir, sitios que
#'
#' ## Elegir método y número de clústers
#'
#' Existen varios criterios para elegir un dendrograma idóneo, como por ejemplo, los gráficos ti
#'
#' ### Seleccionar método de agrupamiento por correlación cofenética
#'
#' La correlación cofenética implica conocer la distancia cofenética, y esta última se entiende m
#'
#' Usando la lista de objetos de clústers, calcularé la correlación cofenética dentro de un 'map
#'
map_df(lista_cl, function(x) {
  coph_d <- cophenetic(x)
  corr <- cor(mi_fam_norm_d, coph_d)
  return(corr)
})

```

```

})
#’
#’ Habrás notado que, tanto UPGMA como enlace completo, tienen valores altos de correlación cofe
#’
#’ ### Elegir número de clústers
#’
#’ Elegiré UPGMA como método de agrupamiento y determinaré cuántos grupos son idóneos de acuerdo
#’
#’ El objetivo de la función ‘calcular_anchuras_siluetas’ está implícito en su nombre, y requiere
#’
#’ 1. Las anchuras promedio para cada partición, excepto para las particiones ‘i=1’ y ‘i=50’, po
#’
#’ 2. Número óptimo de grupos. Haré los cálculos para UPGMA y Ward, y luego explico en qué consi
#’
#’ Para UPGMA:
#’
anch_sil_upgma <- calcular_anchuras_siluetas(
  mc_orig = mi_fam,
  distancias = mi_fam_norm_d,
  cluster = lista_cl$cl_upgma)
anch_sil_upgma
#’
#’ El objeto ‘anchuras_siluetas’ de la lista ‘anch_sil_upgma’ te muestra un vector con los prome
#’
#’ Igualmente, el objeto ‘n_grupos_optimo’ te indica cuál es el número óptimo de clústers a crea
#’
#’ Haré el gráfico de dendrograma, aunque nota que en este caso primero reordenaré los sitios co
#’
u_dend_reord <- reorder.hclust(lista_cl$cl_upgma, mi_fam_norm_d)
plot(u_dend_reord, hang = -1)
rect.hclust(
  tree = u_dend_reord,
  k = anch_sil_upgma$n_grupos_optimo)
#’
#’ Ahora compararé el dendrograma con el mapa de calor en un mismo gráfico, colocando los dendro
#’
heatmap(
  as.matrix(mi_fam_norm_d),
  Rowv = as.dendrogram(u_dend_reord),
  symm = TRUE,
  margin = c(3, 3),
  col = rev(cm.colors(4))
)
#’
#’ En general, hay dos grupos, uno grande y otro pequeño, y parece haber un tercero en el mapa d
#’
#’ Mostraré el resultado para Ward:

```



```

#’
anch_sil_ward <- calcular_anchuras_siluetas(
  mc_orig = mi_fam,
  distancias = mi_fam_norm_d,
  cluster = lista_cl$cl_ward)
anch_sil_ward
#’
#’ En este caso, el valor máximo, que ocupa la posición número 2, no se diferencia mucho del de
#’
w_dend_reord <- reorder.hclust(lista_cl$cl_ward, mi_fam_norm_d)
plot(w_dend_reord, hang = -1)
rect.hclust(
  tree = w_dend_reord,
  k = anch_sil_ward$n_grupos_optimo)
plot(w_dend_reord, hang = -1)
rect.hclust(
  tree = w_dend_reord,
  k = anch_sil_ward$n_grupos_optimo + 1)
#’
#’ Comparando el dendrograma con el mapa de calor. Verificar si el número de grupos hace sentido
#’
heatmap(
  as.matrix(mi_fam_norm_d),
  Rowv = as.dendrogram(w_dend_reord),
  symm = TRUE,
  margin = c(3, 3),
  col = rev(cm.colors(4))
)
#’
#’ Nótese que este dendrograma hace más sentido que el sugerido por UPGMA. En cualquier casos, c
#’
#’ ### Evaluación mediante remuestreo por *bootstrap* multiescalar
#’
#’ Con suerte, un agrupamiento aplicado a datos muestrales reflejará los patrones naturales de o
#’
#’ Sin embargo, los datos de BCI también tienen sesgo, pues se usa un DAP de corte para decidir
#’
#’ No obstante, aun con todas sus bondades, los datos censales carecen de una fortaleza: no refl
#’
#’ Por estas razones, los análisis de agrupamientos realizados hasta este punto, reflejan tanto
#’
#’ El método de remuestreo por *bootstrap* multiescalar está implementado en el paquete ‘pvclust’
#’
#’ La función ‘pvclust’ devolverá un dendrograma enriquecido, que incluirá los valores de AU y BP
#’
#’ - Rectángulos de borde azul, para todos aquellos grupos que resulten con valores de AU>0.91 e
#’

```

```

#’ - Líneas inferiores rojas, que resaltan aquellos grupos (o subgrupos) que obtuvieron AU>0.95.
#’
#’ Ten presente que, al realizar remuestreo por *bootstrap* multiescalar, cada corrida puede arrojar
#’
#’ ##### UPGMA
#’
cl_pvclust_upgma <-
  pvclust(t(mi_fam_norm),
    method.hclust = "average",
    method.dist = "euc",
    iseed = 91, # Resultado reproducible
    parallel = TRUE)
# Añadir los valores de p
plot(cl_pvclust_upgma, hang = -1)
# Añadir rectángulos a los grupos significativos
lines(cl_pvclust_upgma)
pvrect(cl_pvclust_upgma, alpha = 0.91, border = 4)
#’
#’ ##### Ward
#’
cl_pvclust_ward <-
  pvclust(t(mi_fam_norm),
    method.hclust = "ward.D2",
    method.dist = "euc",
    iseed = 191, # Resultado reproducible
    parallel = TRUE)
# Añadir los valores de p
plot(cl_pvclust_ward, hang = -1)
# Añadir rectángulos a los grupos significativos
lines(cl_pvclust_ward)
pvrect(cl_pvclust_ward, alpha = 0.91, border = 4)
#’
#’ ### Recapitulando los grupos de sitios.
#’
#’ ##### Patrones comunes y dispares
#’
#’ Detecto algunos patrones consistentes en cuanto a grupos de sitios según composición de las especies.
#’
#’ - Tanto en UPGMA como en Ward, detecté al menos dos o tres grandes grupos. Con el primer método
#’
#’ - En el caso específico del dendrograma Ward, ‘pvclust’ atomizó los sitios en demasiados grupos.
#’
#’ ##### ¿Cómo declaro los grupos de sitios?
#’
#’ Para conservar las clasificaciones de grupos de sitios anteriores, crearé un vector con el id de cada sitio.
#’
#’ UPGMA:

```

```

(grupos_upgma_k2 <- as.factor(cutree(lista_cl$cl_upgma, k = 2)))
#'
#' En este caso, los sitios 1 y 2 pertenecen al grupo 1, los sitios 3 al 6 pertenecen al grupo 2
#'
table(grupos_upgma_k2)
#'
#' Nota lo desiguales que son estos grupos, un efecto esperado dado el alto grado de autocorrelación
#'
#' Ward:
#'
(grupos_ward_k4 <- as.factor(cutree(lista_cl$cl_ward, k = 4)))
table(grupos_ward_k4)
#'
#' Guardaré estos vectores en archivos para reutilizarlos en *scripts* posteriores:
#'
saveRDS(grupos_upgma_k2, 'grupos_upgma_k2.RDS')
saveRDS(grupos_ward_k4, 'grupos_ward_k4.RDS')
#'
#' Evita usar este, y cualquier otro procedimiento, de manera mecánica. En tu caso, quizá tengas
#'
#' En el próximo *script*, aprenderás a comparar este resultado con las variables ambientales. T

```

aa_3.R

```

#' ---
#' title: "Análisis de agrupamiento (cluster analysis). <br> Parte 3: Grupos (clústers), variables ambientales"
#' author: "JR"
#' date: "15 de noviembre, 2020"
#' output: github_document
#' ---

knitr::opts_chunk$set(fig.width=12, fig.height=8)

#' ## Preámbulo
#'
#' ### Cargar paquetes
#'
library(mapview)
library(tidyverse)
library(sf)
library(RColorBrewer)
source('biodata/funciones.R')
#'
#' ### Cargar datos
#'

```

```

load('biodata/Myrtaceae.Rdata')
load('biodata/matriz_ambiental.Rdata')
grupos_upgma_k2 <- readRDS('grupos_upgma_k2.RDS')
table(grupos_upgma_k2) #Importante, tener en cuenta los desiguales tamaños de los grupos
grupos_ward_k4 <- readRDS('grupos_ward_k4.RDS')
table(grupos_ward_k4)
#'
## Paletas
#'
rojo <- colorRampPalette(brewer.pal(8, "Reds"))
rojo_inv <- colorRampPalette(rev(brewer.pal(8, "Reds")))
colores_grupos <- brewer.pal(8, "Accent")
azul <- colorRampPalette(brewer.pal(8, "Blues"))
azul_inv <- colorRampPalette(rev(brewer.pal(8, "Blues")))
verde <- colorRampPalette(brewer.pal(8, "Greens"))
verde_inv <- colorRampPalette(rev(brewer.pal(8, "Greens")))
gris <- colorRampPalette(brewer.pal(8, "Greys"))
gris_inv <- colorRampPalette(rev(brewer.pal(8, "Greys")))
naranja <- colorRampPalette(brewer.pal(8, "Oranges"))
naranja_inv <- colorRampPalette(rev(brewer.pal(8, "Oranges")))
pastel1 <- colorRampPalette(brewer.pal(8, "YlOrBr"))
pastel1_inv <- colorRampPalette(rev(brewer.pal(8, "YlOrBr")))
pastel2 <- colorRampPalette(brewer.pal(8, "YlOrRd"))
pastel2_inv <- colorRampPalette(rev(brewer.pal(8, "YlOrRd")))
PRGn <- colorRampPalette(brewer.pal(8, "PRGn"))
PRGn_inv <- colorRampPalette(rev(brewer.pal(8, "PRGn")))
PuBu <- colorRampPalette(brewer.pal(8, "PuBu"))
PuBu_inv <- colorRampPalette(rev(brewer.pal(8, "PuBu")))
#'
## Explorar efectos
#'
### Pruebas de igualdad de promedios de las variables entre 2 grupos
#'
# Para evaluar homogeneidad de promedios usaré las pruebas *t* (medias), basada en la distribución
# Primero crearé un objeto que permita realizar tanto las pruebas como los diagramas de cajas.
#'
(m_amb_upgma_k2 <- bci_env_grid %>%
  select_if(is.numeric) %>% select(-id) %>%
  mutate(grupos_upgma_k2) %>%
  st_drop_geometry() %>%
  pivot_longer(-grupos_upgma_k2, names_to = "variable", values_to = "valor"))
#'
# A continuación, las pruebas:
#'
m_amb_upgma_k2 %>%
  group_by(variable) %>%

```

```

summarise(
  p_valor_t = t.test(valor ~ grupos_upgma_k2)$p.value,
  p_valor_w = wilcox.test(valor ~ grupos_upgma_k2, exact = F)$p.value) %>%
arrange(p_valor_t) %>%
print(n=Inf)
#'
#' Interesa observar las variables que obtuvieron valores de  $p < 0.01$ . Reitero que, en mi caso, mi
#'
#' Por otra parte, este es un buen momento para "revisitar" tus análisis exploratorios de datos
#'
#' Los gráficos:
#'
m_amb_upgma_k2 %>%
  group_by(variable) %>%
  ggplot() + aes(x = grupos_upgma_k2, y = valor, fill = grupos_upgma_k2) +
  geom_boxplot() +
  scale_fill_brewer(palette = 'Accent') +
  theme_bw() +
  theme(legend.position="none") +
  facet_wrap(~ variable, scales = 'free_y')
#'
#' Mapas:
#'
mapa_upgma_k2 <- mapView(
  bci_env_grid %>% mutate(grupos_upgma_k2),
  layer.name = 'Grupos (2) UPGMA',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T,
  col.regions = colores_grupos[1:2],
  zcol = 'grupos_upgma_k2') %>%
  addStaticLabels(label = bci_env_grid$id) %>%
  leaflet::setView(
    lng = -79.85136,
    lat = 9.15097,
    zoom = 15)
mapa_upgma_k2
mapa_upgma_k2 %>% mapshot(
  file = 'mapa_upgma_k2.png',
  remove_controls = c("zoomControl", "layersControl", "homeButton")
)
#'
#' Mapa de una de las variables donde se presentó efecto de su promedio ( $p < 0.01$ ), en este caso,
#'
mapa_zn <- mapView(
  bci_env_grid,
  layer.name = 'Zinc',

```

```

alpha.regions = 0.6,
map.types = 'OpenTopoMap',
legend = T,
col.regions = rojo,
zcol = 'Zn') %>%
addStaticLabels(label = bci_env_grid$id) %>%
leaflet::setView(
  lng = -79.85136,
  lat = 9.15097,
  zoom = 15)
mapa_zn
mapa_zn %>% mapshot(
  file = 'mapa_zinc.png',
  remove_controls = c("zoomControl", "layersControl", "homeButton")
)
#'
#' ### Pruebas de igualdad de promedios de las variables entre 3 grupos o más
#'
#' Objeto común:
#'
(m_amb_ward_k4 <- bci_env_grid %>%
  select_if(is.numeric) %>% select(-id) %>%
  mutate(grupos_ward_k4) %>%
  st_drop_geometry() %>%
  pivot_longer(grupos_ward_k4, names_to = "variable", values_to = "valor"))
#'
#' Pruebas, en este caso ANOVA (evalúa homogeneidad de medias; no se cumplen muchos de los supue
#'
m_amb_ward_k4 %>%
  group_by(variable) %>%
  summarise(
    p_valor_a = oneway.test(valor ~ grupos_ward_k4)$p.value,
    p_valor_k = kruskal.test(valor ~ grupos_ward_k4)$p.value) %>%
  arrange(p_valor_k) %>%
  print(n=Inf)
#'
#' Gráficos:
#'
m_amb_ward_k4 %>%
  group_by(variable) %>%
  ggplot() + aes(x = grupos_ward_k4, y = valor, fill = grupos_ward_k4) +
  geom_boxplot() +
  scale_fill_brewer(palette = 'Accent') +
  theme_bw() +
  theme(legend.position="none") +
  facet_wrap(~ variable, scales = 'free_y')
#'

```

```

#' Mapas:
#'
mapa_ward_k4 <- mapView(
  bci_env_grid %>% mutate(grupos_ward_k4),
  layer.name = 'Grupos (4) Ward',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T,
  col.regions = colores_grupos[1:4],
  zcol = 'grupos_ward_k4') %>%
  addStaticLabels(label = bci_env_grid$id) %>%
  leaflet::setView(
    lng = -79.85136,
    lat = 9.15097,
    zoom = 15)
mapa_ward_k4
mapa_ward_k4 %>% mapshot(
  file = 'mapa_ward_k4.png',
  remove_controls = c("zoomControl", "layersControl", "homeButton")
)
#'
#' Mapa de una de las variables donde se presentó efecto de su promedio ( $p < 0.01$ ), en este caso,
#'
mapa_abundanciagl <- mapView(
  bci_env_grid,
  layer.name = 'abundancia_global',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T,
  col.regions = azul_inv,
  zcol = 'abundancia_global') %>%
  addStaticLabels(label = bci_env_grid$id) %>%
  leaflet::setView(
    lng = -79.85136,
    lat = 9.15097,
    zoom = 15)
mapa_abundanciagl
mapa_abundanciagl %>% mapshot(
  file = 'mapa_abundanciagl.png',
  remove_controls = c("zoomControl", "layersControl", "homeButton")
)
#'
mapa_curvatura <- mapView(
  bci_env_grid,
  layer.name = 'curvatura_perfil_media',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',

```

```

legend = T,
col.regions = rojo_inv,
zcol = 'curvatura_perfil_media') %>%
addStaticLabels(label = bci_env_grid$id) %>%
leaflet::setView(
  lng = -79.85136,
  lat = 9.15097,
  zoom = 15)
mapa_curvatura
mapa_curvatura %>% mapshot(
  file = 'mapa_curvatura.png',
  remove_controls = c("zoomControl", "layersControl", "homeButton")
)
#'
mapa_A1 <- mapView(
  bci_env_grid,
  layer.name = 'A1',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T,
  col.regions = azul_inv,
  zcol = 'A1') %>%
addStaticLabels(label = bci_env_grid$id) %>%
leaflet::setView(
  lng = -79.85136,
  lat = 9.15097,
  zoom = 15)
mapa_A1
mapa_A1 %>% mapshot(
  file = 'mapa_A1.png',
  remove_controls = c("zoomControl", "layersControl", "homeButton")
)
#'
mapa_Fe <- mapView(
  bci_env_grid,
  layer.name = 'Fe',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T,
  col.regions = verde_inv,
  zcol = 'Fe') %>%
addStaticLabels(label = bci_env_grid$id) %>%
leaflet::setView(
  lng = -79.85136,
  lat = 9.15097,
  zoom = 15)
mapa_Fe

```



```

mapa_Fe %>% mapshot(
  file = 'mapa_Fe.png',
  remove_controls = c("zoomControl", "layersControl", "homeButton")
)
#'
mapa_curva <- mapView(
  bci_env_grid,
  layer.name = 'curvatura_tangencial_media',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T,
  col.regions = gris_inv,
  zcol = 'curvatura_tangencial_media') %>%
  addStaticLabels(label = bci_env_grid$id) %>%
  leaflet::setView(
    lng = -79.85136,
    lat = 9.15097,
    zoom = 15)
mapa_curva
mapa_curva %>% mapshot(
  file = 'mapa_curva.png',
  remove_controls = c("zoomControl", "layersControl", "homeButton")
)
#'
mapa_elevacion <- mapView(
  bci_env_grid,
  layer.name = 'elevacion_media',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T,
  col.regions = naranja_inv,
  zcol = 'elevacion_media') %>%
  addStaticLabels(label = bci_env_grid$id) %>%
  leaflet::setView(
    lng = -79.85136,
    lat = 9.15097,
    zoom = 15)
mapa_elevacion
mapa_elevacion %>% mapshot(
  file = 'mapa_elevacion.png',
  remove_controls = c("zoomControl", "layersControl", "homeButton")
)
#'
mapa_Mn <- mapView(
  bci_env_grid,
  layer.name = 'Mn',
  alpha.regions = 0.6,

```

```

map.types = 'OpenTopoMap',
legend = T,
col.regions = pastel1_inv,
zcol = 'Mn') %>%
addStaticLabels(label = bci_env_grid$id) %>%
leaflet::setView(
  lng = -79.85136,
  lat = 9.15097,
  zoom = 15)
mapa_Mn
mapa_Mn %>% mapshot(
  file = 'mapa_Mn.png',
  remove_controls = c("zoomControl", "layersControl", "homeButton")
)
#'
mapa_N.min. <- mapView(
  bci_env_grid,
  layer.name = 'N.min.',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T,
  col.regions = pastel2_inv,
  zcol = 'N.min.') %>%
addStaticLabels(label = bci_env_grid$id) %>%
leaflet::setView(
  lng = -79.85136,
  lat = 9.15097,
  zoom = 15)
mapa_N.min.
mapa_N.min. %>% mapshot(
  file = 'mapa_Nin.min.png',
  remove_controls = c("zoomControl", "layersControl", "homeButton")
)
#'
mapa_P <- mapView(
  bci_env_grid,
  layer.name = 'P',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T,
  col.regions = PuBu_inv,
  zcol = 'P') %>%
addStaticLabels(label = bci_env_grid$id) %>%
leaflet::setView(
  lng = -79.85136,
  lat = 9.15097,
  zoom = 15)

```

```

mapa_P
mapa_P %>% mapshot(
  file = 'mapa_P.png',
  remove_controls = c("zoomControl", "layersControl", "homeButton")
)
#'
mapa_heterogeneidadambiental <- mapView(
  bci_env_grid,
  layer.name = 'heterogeineidad_ambiental',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T,
  col.regions = PuBu_inv,
  zcol = 'heterogeneidad_ambiental') %>%
  addStaticLabels(label = bci_env_grid$id) %>%
  leaflet::setView(
    lng = -79.85136,
    lat = 9.15097,
    zoom = 15)
mapa_heterogeneidadambiental
mapa_heterogeneidadambiental %>% mapshot(
  file = 'mapa_heterogeneidad.png',
  remove_controls = c("zoomControl", "layersControl", "homeButton")
)
#'
mapa_orientacionmedia <- mapView(
  bci_env_grid,
  layer.name = 'orientacion_media',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T,
  col.regions = PuBu_inv,
  zcol = 'orientacion_media') %>%
  addStaticLabels(label = bci_env_grid$id) %>%
  leaflet::setView(
    lng = -79.85136,
    lat = 9.15097,
    zoom = 15)
mapa_orientacionmedia
mapa_orientacionmedia %>% mapshot(
  file = 'mapa_orientacion.png',
  remove_controls = c("zoomControl", "layersControl", "homeButton")
)
#'
mapa_riquezaglobal <- mapView(
  bci_env_grid,
  layer.name = 'riqueza_global',

```

```

alpha.regions = 0.6,
map.types = 'OpenTopoMap',
legend = T,
col.regions = PuBu_inv,
zcol = 'riqueza_global') %>%
addStaticLabels(label = bci_env_grid$id) %>%
leaflet::setView(
  lng = -79.85136,
  lat = 9.15097,
  zoom = 15)
mapa_riquezaglobal
mapa_riquezaglobal %>% mapshot(
  file = 'mapa_riquezaglobal.png',
  remove_controls = c("zoomControl", "layersControl", "homeButton")
)
#'
mapa_geohombrera <- mapView(
  bci_env_grid,
  layer.name = 'geomorf_hombrera_pct',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T,
  col.regions = azul_inv,
  zcol = 'geomorf_hombrera_pct') %>%
addStaticLabels(label = bci_env_grid$id) %>%
leaflet::setView(
  lng = -79.85136,
  lat = 9.15097,
  zoom = 15)
mapa_geohombrera
mapa_geohombrera %>% mapshot(
  file = 'mapa_geohombrera.png',
  remove_controls = c("zoomControl", "layersControl", "homeButton")
)
#'
mapa_Cu <- mapView(
  bci_env_grid,
  layer.name = 'Cu',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T,
  col.regions = azul_inv,
  zcol = 'Cu') %>%
addStaticLabels(label = bci_env_grid$id) %>%
leaflet::setView(
  lng = -79.85136,
  lat = 9.15097,

```

```

    zoom = 15)
mapa_Cu
mapa_Cu %>% mapshot(
  file = 'mapa_Cu.png',
  remove_controls = c("zoomControl", "layersControl", "homeButton")
)
#'
mapa_utm_ns <- mapView(
  bci_env_grid,
  layer.name = 'UTM.NS',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T,
  col.regions = azul_inv,
  zcol = 'UTM.NS') %>%
  addStaticLabels(label = bci_env_grid$id) %>%
  leaflet::setView(
    lng = -79.85136,
    lat = 9.15097,
    zoom = 15)
mapa_utm_ns
mapa_utm_ns %>% mapshot(
  file = 'mapa_utm_ns.png',
  remove_controls = c("zoomControl", "layersControl", "homeButton")
)
#'

```

aa_4.R

```

#' ---
#' title: "Análisis de agrupamiento (cluster analysis). <br> Parte 4: Especies indicadoras, espe
#' author: "JR"
#' date: "15 de noviembre, 2020"
#' output: github_document
#' ---

knitr::opts_chunk$set(fig.width=12, fig.height=8)

#' ## Preámbulo
#'
#' ### Cargar paquetes
#'
library(indicspecies)
source('biodata/funciones.R')
#'

```

```

#' ### Cargar datos
#'
load('biodata/Myrtaceae.Rdata')
mi_fam <- mc_myrtc
grupos_upgma_k2 <- readRDS('grupos_upgma_k2.RDS')
table(grupos_upgma_k2)
grupos_ward_k4 <- readRDS('grupos_ward_k4.RDS')
table(grupos_ward_k4)
#'
#' ## Análisis de especies indicadoras mediante IndVal
#'
#' ### UPGMA
#'
iva_upgma_k2 <- multipatt(
  x = mi_fam,
  cluster = grupos_upgma_k2,
  func = 'IndVal.g',
  max.order = 1,
  control = how(nperm = 999))
summary(iva_upgma_k2, indvalcomp = TRUE)
colSums(mi_fam)
(p_upgma_adj <- p.adjust(iva_upgma_k2$sign$p.value))
(iva_upgma_boot <- strassoc(
  X = mi_fam,
  cluster = grupos_upgma_k2,
  func = "IndVal.g",
  nboot = 1000))
#'
#' Ward
#'
iva_ward_k4 <- multipatt(
  x = mi_fam,
  cluster = grupos_ward_k4,
  func = 'IndVal.g',
  max.order = 2,
  control = how(nperm = 999))
summary(iva_ward_k4, indvalcomp = TRUE)
colSums(mi_fam)
(p_ward_adj <- p.adjust(iva_ward_k4$sign$p.value))
(iva_ward_boot <- strassoc(
  X = mi_fam,
  cluster = grupos_ward_k4,
  func = "IndVal.g",
  nboot = 1000))
#'
#' ## Análisis de especies con preferencia por hábitat mediante el coeficiente de correlación bi
#'

```

```

#' ### UPGMA
#'
phi_upgma_k2 <- multipatt(
  mi_fam,
  grupos_upgma_k2,
  func = "r.g",
  max.order = 1,
  control = how(nperm = 999))
summary(phi_upgma_k2)
colSums(mi_fam)
(phi_upgma_boot <- strassoc(
  X = mi_fam,
  cluster = grupos_upgma_k2,
  func = "r.g",
  nboot = 1000))
#'
#' Ward
#'
phi_ward_k4 <- multipatt(
  mi_fam,
  grupos_ward_k4,
  func = "r.g",
  max.order = 2,
  control = how(nperm = 999))
summary(phi_ward_k4)
colSums(mi_fam)
(phi_ward_boot <- strassoc(
  X = mi_fam,
  cluster = grupos_ward_k4,
  func = "r.g",
  nboot = 1000))

```

to_1.R

```

#' ---
#' title: "Técnicas de ordenación. <br> Parte 1: Ordenación no restringida. <br> PCA, CA y PCoA"
#' author: "JR"
#' date: "21 de noviembre, 2020"
#' output: github_document
#'
#' ---

knitr::opts_chunk$set(fig.width=12, fig.height=8)

#' ## Preámbulo

```

```

#'
#' ### Cargar paquetes
#'
library(vegan)
library(tidyverse)
library(sf)
library(mapview)
source('biodata/funciones.R')
#'
#' ### Cargar datos
#'
load('biodata/Myrtaceae.Rdata')
load('biodata/matriz_ambiental.Rdata')
mi_fam <- mc_myrtc
(colnames(mi_fam) <- make.cepnames(colnames(mi_fam)))
(df_equivalecias <- data.frame(
  nombre_original = colnames(mc_myrtc),
  colnames(mi_fam)))
bci_env_grid %>% tibble
grupos_upgma_k2 <- readRDS('grupos_upgma_k2.RDS')
table(grupos_upgma_k2)
grupos_ward_k4 <- readRDS('grupos_ward_k4.RDS')
table(grupos_ward_k4)
#'
#' ## Ordenación
#'
#' La ordenación se basa en los mismos principios que la medición de asociación (similaridad) y
#'
#' A diferencia del análisis de agrupamiento, o como complemento de éste, el análisis de ordenación
#'
#' El análisis de ordenación puede ser no restringido (o simple) y restringido (o 'canónico'). E
#'
#' Las principales técnicas de ordenación no restringida son análisis de componentes principales
#'
#' ### Análisis de componentes principales (PCA)
#'
#' Es el método tradicional basado en vectores propios que comúnmente se aplica a datos cuantitativos
#'
#' #### PCA aplicado a datos ambientales
#'
#' Para aplicar PCA a datos ambientales, es necesario que todas las variables sean numéricas y
#'
#' Dado que se requiere que las variables de entrada sean exclusivamente numéricas, el primer paso es
#'
#' ¡IMPORTANTE! Haré esta demostración sólo con las variables de suelo, **pero puedes (y debes)
#'
#' A partir de los datos de suelo, la función 'rda', de 'vegan' realizará los siguientes pasos:

```



```

# '
env_suelo <- bci_env_grid %>%
  st_drop_geometry %>%
  dplyr::select(matches('^([A-T,Z]|~pH$'), ignore.case = F))
env_suelo %>% tibble
env_suelo_pca <- rda(env_suelo, scale = TRUE)
env_suelo_pca
summary(env_suelo_pca)
# '
env_geomorf <- bci_env_grid %>%
  select(curvatura_perfil_media, curvatura_tangencial_media, elevacion_media, geomorf_hombrera_p,
  st_drop_geometry()
env_geomorf %>% tibble
env_geomorf_pca <- rda(env_geomorf, scale = TRUE)
env_geomorf_pca
summary(env_geomorf_pca)
# ' Para agilizar la producción de scripts analíticos de referencia, trasladaré las explicaciones
# '
# ' En el vídeo asociado, explico el significado de:
# '
# ' - Inercia, *Inertia*
# ' - Valores propios, autovalores, *Eigenvalues*
# ' - Escalamiento, *Scaling*
# ' - Puntuaciones de "especies", *Species scores*
# ' - Puntuaciones de "sitios", *Site scores*
# '
screepplot(env_suelo_pca, bstick = TRUE)
screepplot(env_geomorf_pca, bstick = TRUE)
# '
# ' Usando función 'cleanplot.pca'
# '
par(mfrow = c(1, 2))
cleanplot.pca(env_suelo_pca, scaling = 1, mar.percent = 0.08, cex.char1 = 0.8)
cleanplot.pca(env_suelo_pca, scaling = 2, mar.percent = 0.04, cex.char1 = 0.8)
par(mfrow = c(1, 1))
# '
par(mfrow = c(1, 2))
cleanplot.pca(env_geomorf_pca, scaling = 1, mar.percent = 0.08, cex.char1 = 0.8)
cleanplot.pca(env_geomorf_pca, scaling = 2, mar.percent = 0.04, cex.char1 = 0.8)
par(mfrow = c(1, 1))
# ' Comparar distribución de los sitios en biplots con distribución real en el mapa:
# '
# ' ### Generar mapa de cuadros sin simbología
# '
mapa_cuadros <- mapView(
  bci_env_grid,
  col.regions = 'grey80',

```

```

alpha.regions = 0.3,
map.types = 'OpenTopoMap',
legend = F, zoom = 14,
zcol = 'id') %>% addStaticLabels() %>%
leaflet::setView(
  lng = -79.85136,
  lat = 9.15097,
  zoom = 15)
mapa_cuadros
#'
#' Comparar con resultados de un análisis de agrupamiento del mismo conjunto de datos. Primero a
#'
(env_agrupamiento <- hclust(dist(scale(env_suelo)), 'ward.D'))
(env_grupos <- cutree(env_agrupamiento, k = 3))
(mi_cluster <- factor(env_grupos))
(mi_cluster_l <- levels(mi_cluster))
(mi_cluster_l_seq <- 1:length(mi_cluster_l))
#'
(env_agrupamiento <- hclust(dist(scale(env_geomorf)), 'ward.D'))
(env_grupos <- cutree(env_agrupamiento, k = 4))
(mi_cluster <- factor(env_grupos))
(mi_cluster_l <- levels(mi_cluster))
(mi_cluster_l_seq <- 1:length(mi_cluster_l))
#' Observa que estoy generando un agrupamiento basado en los datos de suelo. No estoy comparando
#'
#' Luego calculo las puntuaciones de los sitios para usarlas luego como coordenadas de los puntos
#'
(puntuaciones <- scores(env_suelo_pca, display = 'wa', scaling = 1))
(puntuaciones <- scores(env_geomorf_pca, display = 'wa', scaling = 1))
#'
#' Luego creo el gráfico base, coloco los puntos sobre el gráfico usando las puntuaciones, les c
#'
grafico_base <- plot(
  env_suelo_pca,
  display = "wa",
  scaling = 1,
  type = "n",
  main = "PCA y grupos"
)
abline(v = 0, lty = "dotted")
abline(h = 0, lty = "dotted")
for (i in mi_cluster_l_seq) {
  points(puntuaciones[mi_cluster == i, ],
    pch = (14 + i),
    cex = 2,
    col = i + 1)
}

```

```

text(puntuaciones, row.names(env_suelo), cex = 1, pos = 3)
legend(
  "topright", # Otras alternativas: "bottomleft", "bottomright" y "topleft"
  paste("Grupo", c(mi_cluster_l_seq)),
  pch = 14 + c(mi_cluster_l_seq),
  col = 1 + c(mi_cluster_l_seq),
  pt.cex = 2
)
# '
grafico_base <- plot(
  env_geomorf_pca,
  display = "wa",
  scaling = 1,
  type = "n",
  main = "PCA y grupos"
)
abline(v = 0, lty = "dotted")
abline(h = 0, lty = "dotted")
for (i in mi_cluster_l_seq) {
  points(puntuaciones[mi_cluster == i, ],
    pch = (14 + i),
    cex = 2,
    col = i + 1)
}
text(puntuaciones, row.names(env_geomorf), cex = 1, pos = 3)
legend(
  "topright", # Otras alternativas: "bottomleft", "bottomright" y "topleft"
  paste("Grupo", c(mi_cluster_l_seq)),
  pch = 14 + c(mi_cluster_l_seq),
  col = 1 + c(mi_cluster_l_seq),
  pt.cex = 2
)
# ' Es razonable que el análisis cluster y el biplot muestren patrones consistentes, puesto que s
# '
# ' Si hago lo mismo, pero usando mi análisis de agrupamiento anterior (*scripts* "aa_analisis_de
# '
(mi_cluster_anterior <- grupos_upgma_k2)
(mi_cluster_anterior <- grupos_ward_k4)
(mi_cluster_anterior_l <- levels(mi_cluster_anterior))
(mi_cluster_anterior_l_seq <- 1:length(mi_cluster_anterior_l))
grafico_base <- plot(
  env_suelo_pca,
  display = "wa",
  scaling = 1,
  type = "n",
  main = "PCA y grupos"
)

```

```

abline(v = 0, lty = "dotted")
abline(h = 0, lty = "dotted")
for (i in mi_cluster_anterior_l_seq) {
  points(puntuaciones[mi_cluster_anterior == i, ],
        pch = (14 + i),
        cex = 2,
        col = i + 1)
}
text(puntuaciones, row.names(env_suelo), cex = 1, pos = 3)
legend(
  "topright", # Otras alternativas: "bottomleft", "bottomright" y "topleft"
  paste("Grupo", c(mi_cluster_anterior_l_seq)),
  pch = 14 + c(mi_cluster_anterior_l_seq),
  col = 1 + c(mi_cluster_anterior_l_seq),
  pt.cex = 2
)
# '
(mi_cluster_anterior_l <- levels(mi_cluster_anterior))
(mi_cluster_anterior_l_seq <- 1:length(mi_cluster_anterior_l))
grafico_base <- plot(
  env_geomorf_pca,
  display = "wa",
  scaling = 1,
  type = "n",
  main = "PCA y grupos"
)
abline(v = 0, lty = "dotted")
abline(h = 0, lty = "dotted")
for (i in mi_cluster_anterior_l_seq) {
  points(puntuaciones[mi_cluster_anterior == i, ],
        pch = (14 + i),
        cex = 2,
        col = i + 1)
}
text(puntuaciones, row.names(env_geomorf), cex = 1, pos = 3)
legend(
  "topright", # Otras alternativas: "bottomleft", "bottomright" y "topleft"
  paste("Grupo", c(mi_cluster_anterior_l_seq)),
  pch = 14 + c(mi_cluster_anterior_l_seq),
  col = 1 + c(mi_cluster_anterior_l_seq),
  pt.cex = 2
)
# ' Esto podría significar que las tendencias/patrones de mi matriz de comunidad (cuadros de 1 Ha
# '
# ' #### PCA aplicado a datos de comunidad transformados
# '
mi_fam_hel <- decostand(mi_fam, method = 'hellinger')

```

```

mi_fam_hel %>% tibble
mi_fam_hel_pca <- rda(mi_fam_hel)
summary(mi_fam_hel_pca)
screplot(
  mi_fam_hel_pca,
  bstick = TRUE,
  npcs = length(mi_fam_hel_pca$CA$eig)
)
mi_fam_hel_pca_sc1 <- scores(mi_fam_hel_pca,
                             display = "species", scaling = 1)
mi_fam_hel_pca_sc2 <- scores(mi_fam_hel_pca,
                             display = "species", scaling = 2)

par(mfrow = c(1, 2))
cleanplot.pca(mi_fam_hel_pca, scaling = 1, mar.percent = 0.06, cex.char1 = 0.7)
cleanplot.pca(mi_fam_hel_pca, scaling = 2, mar.percent = 0.06, cex.char1 = 0.7)
par(mfrow = c(1, 1))
#'
#' Si intentáramos realizar el PCA a datos de comunidad no transformados, no recogeríamos apropi
#'
#' Las especies que contribuyen mucho a los ejes 1 y 2 del PCA (aquellas cuyos vectores sobresal
#'
#' Evaluaré el ajuste del PCA de datos de comunidad a datos ambientales, mediante la función 'en
#'
biplot(
  mi_fam_hel_pca,
  main = "PCA, escalamiento 2, ajuste a variables ambientales")
(mi_fam_hel_pca_envfit <- envfit(mi_fam_hel_pca, env_suelo, scaling = 2))
plot(mi_fam_hel_pca_envfit, p.max = 0.05 , col = 3)
#'
#' Comento los resultados en el video asociado. También probaré ajuste con todas las numéricas d
#'
#' NOTA: te recomiendo probar otros métodos de selección de variables, como por ejemplo, usando
#'
env_num <- bci_env_grid %>%
  select_if(is.numeric) %>%
  select(-id) %>%
  st_drop_geometry
(mi_fam_hel_pca_envfit_num <- envfit(mi_fam_hel_pca, env_num, scaling = 2))
biplot(
  mi_fam_hel_pca,
  main = "PCA, escalamiento 2, ajuste a variables ambientales")
plot(mi_fam_hel_pca_envfit_num, p.max = 0.05 , col = 3)
biplot(
  mi_fam_hel_pca,
  main = "PCA, escalamiento 2, ajuste a variables ambientales")
plot(mi_fam_hel_pca_envfit_num, p.max = 0.1 , col = 3)
#'

```

```

# Comento los resultados en el vídeo asociado.
#
# ¿Cuándo o a qué datos aplicar PCA?
#
# - PCA no es especialmente sensible a datos muy desviados de la normalidad.
# - Como toda técnica, PCA tiene limitaciones.
# - Las variables deben ser dimensionalmente homogéneas (unidades comparables o adimensionales)
# - No usar en matriz transpuestas (no hace sentido la covarianza entre objetos).
# - Es posible usar PCA con dato de presencia/ausencia, en cuyo caso, la matriz de comunidad de
# - Las relaciones entre variables se miden por ángulos, no por proximidad de las puntas de los
#
# ### Análisis de correspondencia (CA)
#
mi_fam_ca <- cca(mi_fam)
summary(mi_fam_ca)
summary(mi_fam_ca, scaling = 1)
#
# Screeplot
#
screeplot(mi_fam_ca, bstick = TRUE, npcs = length(mi_fam_ca$CA$eig))
#
# Biplots
#
par(mfrow = c(1, 2))
plot(mi_fam_ca,
     scaling = 1,
     main = "Análisis de correspondencia, escalamiento 1"
)
plot(mi_fam_ca,
     scaling = 2, # Por defecto scaling=2, lo escribo sólo para fines didáticos
     main = "Análisis de correspondencia, escalamiento 2")
par(mfrow = c(1, 1))
#
# Excluyendo especie *Thevetia ahouai*, abreviada como *Thevahou*.
#
mi_fam_ca <- cca(mi_fam[, -grep('Chamschi', colnames(mi_fam))])
summary(mi_fam_ca)
summary(mi_fam_ca, scaling = 1)
screeplot(mi_fam_ca, bstick = TRUE, npcs = length(mi_fam_ca$CA$eig))
par(mfrow = c(1, 2))
plot(mi_fam_ca,
     scaling = 1,
     main = "CA, escalamiento 1, sin Chamchi"
)
plot(mi_fam_ca,
     scaling = 2,
     main = "CA, escalamiento 2, sin Chamchi")

```

```

par(mfrow = c(1, 1))
#’
#’ Análisis de coordenadas principales (PCoA)
#’
#’ Las técnicas de ordenación anteriores preservan la distancia euclídea entre los objetos. Si n
#’
#’ La función que realiza el PCoA en ‘{vegan}’ es ‘cmdscale’ (de *Classical (Metric) Multidimens
#’
mi_fam_d_bray <- vegdist(mi_fam, method = 'bray') # En realidad, 'bray' es la opción por defecto
mi_fam_d_bray_pcoa <- cmdscale(
  mi_fam_d_bray,
  k = (nrow(mi_fam) - 1),
  add = T,
  eig = TRUE)
round(mi_fam_d_bray_pcoa$eig, 2)
round(sum(mi_fam_d_bray_pcoa$eig[mi_fam_d_bray_pcoa$eig<0]),2)
round(sum(mi_fam_d_bray_pcoa$eig[mi_fam_d_bray_pcoa$eig>=0]),2)
ordiplot(scores(mi_fam_d_bray_pcoa, choices = c(1, 2)),
  type = "t",
  main = "PCoA con promedios ponderados de especies")
abline(h = 0, lty = 3)
abline(v = 0, lty = 3)
mi_fam_d_bray_pcoa_wa <- wascores(mi_fam_d_bray_pcoa$points[, 1:2], mi_fam)
text(
  mi_fam_d_bray_pcoa_wa,
  rownames(mi_fam_d_bray_pcoa_wa),
  cex = 0.7, col = "red")
(mi_fam_d_bray_pcoa_env <- envfit(mi_fam_d_bray_pcoa, env_num))
plot(mi_fam_d_bray_pcoa_env, p.max = 0.05, col = 3)

```

to_2.R

```

#’ ---
#’ title: "Técnicas de ordenación. <br> Parte 2: Ordenación restringida o 'canónica'. <br> RDA,"
#’ author: "JR"
#’ date: "21 de noviembre, 2020"
#’ output: github_document
#’ ---

knitr::opts_chunk$set(fig.width=12, fig.height=8)

#’ ## Preámbulo
#’
#’ ### Cargar paquetes
#’

```

```

library(vegan)
library(tidyverse)
library(sf)
source('biodata/funciones.R')
#'
#' ### Cargar datos
#'
load('biodata/Myrtaceae.Rdata')
load('biodata/matriz_ambiental.Rdata')
mi_fam <- mc_myrtc
(colnames(mi_fam) <- make.cepnames(colnames(mi_fam)))
(df_equivencias <- data.frame(
  nombre_original = colnames(mc_myrtc),
  colnames(mi_fam)))
bci_env_grid %>% tibble
#'
#' ## Ordenación restringida
#'
#' Con este conjunto de técnicas de ordenación, las tendencias detectadas en un conjunto de datos
#'
#' Las principales técnicas de ordenación restringida son análisis de redundancia o RDA (siglas
#'
#' ### Análisis de redundancia (RDA)
#'
#' En el análisis de ordenación simple o no restringida, la matriz de datos (de comunidad o ambiental)
#'
#' RDA combina la regresión y el análisis de componentes principales. El RDA es una regresión lineal
#'
#' - Cada variable de respuesta (e.g. matriz de comunidad) es ajustada (por regresión) a la matriz de variables
#'
#' - Se prueba significancia estadística de la relación entre las variables de respuesta y explicativas
#'
#' - Si la prueba resulta significativa, se realiza un PCA de la matriz de valores ajustados. Es decir, se
#'
#' - Luego se usa la matriz de vectores propios canónicos para calcular dos tipos de puntuaciones:
#'
#'   - Usando la matriz de valores ajustados para obtener una ordenación en el espacio de las variables
#'
#'   - Usando la matriz de valores centrados de las variables de respuesta para obtener una ordenación en el
#'
#' Mostraré a continuación ejemplos aplicados a mi familia asignada. Para agilizar, trasladaré los datos a un
#'
#' Nota (de nuevo): recuerda que los datos de BCI están autocorrelacionados espacialmente, por lo que se
#'
'### Ejemplo usando la matriz ambiental de variables suelo:
'
mi_fam_hel <- decostand(mi_fam, method = 'hellinger')

```



```

mi_fam_hel %>% tibble
env_suelo <- bci_env_grid %>%
  st_drop_geometry %>%
  dplyr::select(matches('^([A-T,Z]|~pH$'), ignore.case = F))
env_suelo %>% tibble
mi_fam_hel_rda_suelo <- rda(mi_fam_hel ~ ., env_suelo)
summary(mi_fam_hel_rda_suelo)
#'
#' ¿Qué partes del resumen debes mirar?
#'
#' - La varianza, particionada, de la cual se muestra tanto la inercia (absoluta) como la propor
#'
#' - Los valores propios (*eigenvalues*) y su contribución a la varianza. Existen *eigenvalues*
#'
#' - Las puntuaciones de sitios (*site scores*) y de especies (*species scores*) tienen el mismo
#'
#' - Las puntuaciones restringidas de sitio (en '{vegan}' se usa la etiqueta *Site constraints (
#'
#' - Finalmente, el resumen presenta las puntuaciones para el *biplot*, que más adelante usaré e
#'
#' Tal como comenté arriba, es importante calcular un  $R^2$  insesgado, siendo este más ú
#'
RsquareAdj(mi_fam_hel_rda_suelo)$adj.r.squared
#'
#' Normalmente, el  $R^2$  insesgado es mucho más bajo que el sesgado, porque se relativiz
#'
#' Otro diagnóstico importante es la determinación de los factores de inflación de la varianza (
#'
vif.cca(mi_fam_hel_rda_suelo)
#'
#' Variables con valores por encima de 10 deben ser examinadas y, desde una posición conservador
#'
#' Finalmente, la representación del modelo se realiza en un *triplot*, que es un gráfico enriqu
#'
#' Escalamiento 1:
#'
plot(mi_fam_hel_rda_suelo,
     scaling = 1,
     display = c("sp", "lc", "cn"),
     main = "Triplot de RDA especies ~ var. suelo, escalamiento 1"
)
mi_fam_hel_rda_suelo_sc1 <-
  scores(mi_fam_hel_rda_suelo,
        choices = 1:2,
        scaling = 1,
        display = "sp"
  )

```

```

arrows(0, 0,
       mi_fam_hel_rda_suelo_sc1[, 1] * 0.9,
       mi_fam_hel_rda_suelo_sc1[, 2] * 0.9,
       length = 0,
       lty = 1,
       col = "red"
)
#'
#' Paso la interpretación de este gráfico y la del siguiente al vídeo correspondiente.
#'
#' Escalamiento 2
#'
plot(mi_fam_hel_rda_suelo,
     scaling = 2,
     display = c("sp", "lc", "cn"),
     main = "Triplot de RDA especies ~ var. suelo, escalamiento 2"
)
mi_fam_hel_rda_suelo_sc2 <-
  scores(mi_fam_hel_rda_suelo,
        scaling = 2,
        choices = 1:2,
        display = "sp"
  )
arrows(0, 0,
       mi_fam_hel_rda_suelo_sc2[, 1] * 0.9,
       mi_fam_hel_rda_suelo_sc2[, 2] * 0.9,
       length = 0,
       lty = 1,
       col = "red"
)
#'
#' #### Ejemplo usando las matriz ambiental con variables seleccionadas
#'
#' El RDA anterior mostró que las variables de suelo son útiles para predecir la matriz de comun
#'
#' - Hay mucha colinealidad entre ellas.
#'
#' - No se probó mejorar el modelo añadiendo otras variables, además de las de suelo.
#'
#' Crearé una matriz ambiental con las variables que resultaron significativas en el ajuste *pos
#'
env_selec <- bci_env_grid %>%
  select(
    heterogeneidad_ambiental,
    riqueza_global,
    UTM.NS,
    Al, B, Ca, Cu, Fe, K, Mg, Mn, P, Zn, N, N.min., pH) %>%

```

```

    st_drop_geometry
mi_fam_hel_rda_selec <- rda(mi_fam_hel ~ ., env_selec)
#'
vif.cca(mi_fam_hel_rda_selec)
#'
#' Haré el plot del escalamiento 2 para comprobar gráficamente asociación entre variables sin la
#'
plot(mi_fam_hel_rda_selec,
     scaling = 2,
     display = c("sp", "lc", "cn"),
     main = "Triplot de RDA especies ~ var. selec, escalamiento 2"
)
#'
#' Tal como comenté arriba, variables con valores VIF por encima de 10, deben ser examinadas. En
#'
env_selec2 <- bci_env_grid %>%
  select(
    heterogeneidad_ambiental,
    riqueza_global,
    UTM.NS,
    Al, B, Cu, Fe, K, Mg, Mn, P, N, N.min., pH) %>%
  st_drop_geometry
mi_fam_hel_rda_selec2 <- rda(mi_fam_hel ~ ., env_selec2)
vif.cca(mi_fam_hel_rda_selec2)
plot(mi_fam_hel_rda_selec2,
     scaling = 2,
     display = c("sp", "lc", "cn"),
     main = "Triplot de RDA especies ~ var. selec2, escalamiento 2"
)
#'
#' Nota que las posiciones rotaron, no así la asociación entre la mayoría de las variables. 'B'
#'
env_selec3 <- bci_env_grid %>%
  select(
    heterogeneidad_ambiental,
    riqueza_global,
    UTM.NS,
    Al, Cu, Fe, K, Mg, Mn, P, N, N.min., pH) %>%
  st_drop_geometry
mi_fam_hel_rda_selec3 <- rda(mi_fam_hel ~ ., env_selec3)
vif.cca(mi_fam_hel_rda_selec3)
plot(mi_fam_hel_rda_selec3,
     scaling = 2,
     display = c("sp", "lc", "cn"),
     main = "Triplot de RDA especies ~ var. selec3, escalamiento 2"
)
#'

```

```

#’ Finalmente, la coordenada ‘UTM.EW’ tiene un alto valor VIF, por lo que es preferible excluirla
#’
env_selec4 <- bci_env_grid %>%
  select(
    riqueza_global,
    UTM.NS,
    Al, Cu, Fe, Mg, Mn, P, N, N.min., pH) %>%
  st_drop_geometry
mi_fam_hel_rda_selec4 <- rda(mi_fam_hel ~ ., env_selec4)
vif.cca(mi_fam_hel_rda_selec4)
plot(mi_fam_hel_rda_selec4,
     scaling = 2,
     display = c("sp", "lc", "cn"),
     main = "Triplot de RDA especies ~ var. selec4, escalamiento 2"
)
#’
#’ Las variables ‘N.min.’ y ‘Zn’ ahora presentan valores VIF aceptables. Habrás notado que es to
#’
summary(mi_fam_hel_rda_selec4)
RsquareAdj(mi_fam_hel_rda_selec4)$adj.r.squared
#’
#’ Triplot, pero ahora con las flechas para las especies.
#’
#’ Escalamiento 1:
#’
plot(mi_fam_hel_rda_selec4,
     scaling = 1,
     display = c("sp", "lc", "cn"),
     main = "Triplot de RDA especies ~ var. selec4, escalamiento 1"
)
mi_fam_hel_rda_selec4_sc1 <-
  scores(mi_fam_hel_rda_selec4,
        choices = 1:2,
        scaling = 1,
        display = "sp"
  )
arrows(0, 0,
       mi_fam_hel_rda_selec4_sc1[, 1] * 0.9,
       mi_fam_hel_rda_selec4_sc1[, 2] * 0.9,
       length = 0,
       lty = 1,
       col = "red"
)
#’
#’ Escalamiento 2
#’
plot(mi_fam_hel_rda_selec4,

```

```

    scaling = 2,
    display = c("sp", "lc", "cn"),
    main = "Triplot de RDA especies ~ var. selec4, escalamiento 2"
)
mi_fam_hel_rda_selec4_sc2 <-
  scores(mi_fam_hel_rda_selec4,
    scaling = 2,
    choices = 1:2,
    display = "sp"
  )
arrows(0, 0,
  mi_fam_hel_rda_selec4_sc2[, 1] * 0.9,
  mi_fam_hel_rda_selec4_sc2[, 2] * 0.9,
  length = 0,
  lty = 1,
  col = "red"
)
#'
#' ### Análisis de correspondencia canónica (CCA)
#'
#' #### Ejemplo usando las matriz ambiental con variables seleccionadas:
#'
mi_fam_cca_selec4 <- cca(mi_fam ~ ., env_selec4)
summary(mi_fam_cca_selec4)
RsquareAdj(mi_fam_cca_selec4)
#'
#' Escalamiento 1
#'
plot(mi_fam_cca_selec4,
  scaling = 1,
  display = c("sp", "lc", "cn"),
  main = "Triplot de CCA especies ~ var. selec4, escalamiento 1"
)
#'
#' Escalamiento 2
#'
plot(mi_fam_cca_selec4,
  scaling = 2,
  display = c("sp", "lc", "cn"),
  main = "Triplot de CCA especies ~ var. selec4, escalamiento 2"
)
#'
#' Excluyendo especies con abundancia menor a 100 individuos
#'
colSums(mi_fam) %>% sort
mi_fam_no_raras <- mi_fam %>% select_if(colSums(.) > 100)
intersect(colnames(mi_fam), colnames(mi_fam_no_raras))

```

```

setdiff(colnames(mi_fam), colnames(mi_fam_no_raras))
mi_fam_no_raras_cca_selec4 <- cca(mi_fam_no_raras ~ ., env_selec4)
summary(mi_fam_no_raras_cca_selec4)
RsquareAdj(mi_fam_no_raras_cca_selec4)
#'
#' Escalamiento 1
#'
plot(mi_fam_no_raras_cca_selec4,
     scaling = 1,
     display = c("sp", "lc", "cn"),
     main = "Triplot de CCA especies no raras ~ var. selec4, escalamiento 1"
)
#'
#' Escalamiento 2
#'
plot(mi_fam_no_raras_cca_selec4,
     scaling = 2,
     display = c("sp", "lc", "cn"),
     main = "Triplot de CCA especies no raras ~ var. selec4, escalamiento 2"
)

```

di_1.R

```

#' ---
#' title: "Análisis de diversidad. <br> Parte 1: Diversidad alpha"
#' author: "JR"
#' date: "29 de noviembre, 2020"
#' output: github_document
#' ---

knitr::opts_chunk$set(fig.width=12, fig.height=8)

#' ## Preámbulo
#'
#' ### Cargar paquetes
#'
library(vegan)
library(adespatial)
library(plyr)
library(RColorBrewer)
library(tidyverse)
library(sf)
library(SpadeR)
library(iNEXT)
source('biodata/funciones.R')

```

```

#'
#' ### Cargar datos
#'
load('biodata/Myrtaceae.Rdata')
load('biodata/matriz_ambiental.Rdata')
mi_fam <- mc_myrtc
bci_env_grid %>% tibble
grupos_upgma_k2 <- readRDS('grupos_upgma_k2.RDS')
table(grupos_upgma_k2)
grupos_ward_k4 <- readRDS('grupos_ward_k4.RDS')
table(grupos_ward_k4)
#'
#' ## Diversidad alpha
#'
#' **La tentación de medir la diversidad mediante un único número, ha atrapado a muchos investigadores
#'
#' **La riqueza de especies, una cifra de uso común en estudios ecológicos, es un elemento muy sensible
#'
#' El término "diversidad biológica" se asume que fue acuñado a principios de los 80, y se refiere a la
#'
#' Biodiversidad es de factura más reciente, y es una contracción de "biológica-diversidad"
#'
#' La definición de biodiversidad del PNUMA es: "variabilidad entre organismos vivos" de todos los
#'
#' Hubbell (2001) ofrece una definición más adaptada a la práctica actual y mucho más restrictiva
#'
#' En estas dos últimas acepciones (Hubbell y Magurran), la diversidad biológica puede dividirse en
#'
#' ### La diversidad de especies como un único número
#'
#' Usaré la notación *q* para designar el número de especies o riqueza de especies, y *n* para el número de
#'
#' #### Riqueza de especies y rarefacción
#'
#' Magurran (2004) distingue entre densidad de especies, que equivale al número de especies
#'
#' Para asegurar la comparabilidad entre sitios, se han propuesto distintos métodos. Uno es la rarefacción
#'
#' [](rarefaccion.jpg)
#'
#' #### Componentes de la diversidad de especies basada en abundancia: riqueza y equidad
#'
#' Asumamos que un sitio de muestreo es una variable cualitativa, y cada especie es un "estado".
#'
#' [](shannon.jpg)
#'
#' Desde el punto de vista ecológico, la H tiene dos propiedades importantes:

```

```

#'
#' - Crece al aumentar la riqueza de especies *q*.
#' - Crece con la uniformidad (=equidad o equitabilidad, es decir, qué tan bien repartida se enc
#'
#' Para una *q* dada, *H* asume su valor máximo ( $H_{\max}$ ) cuando todas las especies es
#'
#' [](shannon_max.jpg)
#'
#' Un ejemplo ilustra lo anterior:
#'
#' - Cinco especies, 55 individuos, abundancias desiguales:
#'
foo1 <- c(25, 16, 9, 4, 1)
diversity(foo1)
#'
#' - Cinco especies, 55 individuos, abundancias homogéneas
#'
foo2 <- c(11, 11, 11, 11, 11)
diversity(foo2)
log(5)
#'
#' Otros índices miden exclusivamente la homogeneidad de la abundancia, a lo que se denomina equ
#'
#' Uno de los más usados es la equidad de Pielou o  $J$  (1966), que es la razón entre la entr
#'
#' [](pielou.jpg)
#'
#' Teóricamente, este índice sólo debería aumentar cuando aumenta la homogeneidad de la abundancia
#'
#' Todos los índices de equidad se relacionan con la forma de los modelos de abundancia de esp
#'
#' Otra medida común en trabajos de ecología numérica, pero que es inversa o contraria a la equi
#'
#' [](simpson.jpg)
#'
#' Este valor aumenta con la dominancia, por lo que realmente no mide diversidad, sino más bien
#'
#' La riqueza de especies (*q*), la entropía de Shannon (*H*), la equidad de Pielou ( $J$ ) y el i
#'
#' [](renyi.jpg)
#'
#' donde *a* es el orden de la medida de entropía ( $a=0,1,2,\dots$ ), la cual cuantifica la importan
#'
#' [](hill.jpg)
#'
#' Las tres primeras entropías de Renyi ( $H_a$ ), donde  $a=0, 1$  y  $2$ , y los correspondi
#'

```



```

#' Por otra parte, los tres primeros números de diversidad de Hill tienen significados que cono
#'
#' [](tres_entro_renyi_hill_div_num.jpg) <br>
#' > Según Borcard et al., 2018.
#'
#' Bajo esta notación:
#'
#' - La equidad de Pielou (denominada también equidad de Shannon) equivale a  $J = H_{1\text{st}}/H_{\text{max}}$ 
#'
#' Usando la misma aproximación, Hill propuso también otras ratios que, a diferencia de lo que o
#'
#' -  $E_{1\text{st}} = N_{1\text{st}}/N_{0\text{th}}$  a la cual el propio Hill denominó como su versi
#'
#' -  $E_{2\text{nd}} = N_{2\text{nd}}/N_{0\text{th}}$ , a la cual Hill denominó como su versión de la
#'
#' Los números de diversidad y las ratios de Hill son menos sensibles a las matrices de comunida
#'
#' **Estas afirmaciones tienen implicaciones muy importantes desde el punto de vista ecológico,
#'
#' **Índices, entropías, equidades, ratios**
#'
(indices <- alpha_div(mi_fam))
pairs(indices,
      lower.panel = panel.smooth,
      upper.panel = panel.cor,
      diag.panel = panel.hist,
      main = "Pearson Correlation Matrix")
indices_env <- bind_cols(
  indices,
  bci_env_grid %>%
    select_if(is.numeric) %>%
    st_drop_geometry %>%
    select(-id) %>%
    select(-matches('^geom.*pct$')))
indices_env %>% tibble
ezCorM(indices_env, r_size_lims = c(3,5), label_size = 4)
#'
#' **Modelos de abundancia de especies**
#'
mi_fam_mae <- radfit(mi_fam)
plot(mi_fam_mae)
#'
#' **Rarefacción**
#'
#' Riqueza por sitios
#'
riqueza <- specnumber(mi_fam)

```

```

riqueza %>% sort
#'
#' Sitios con riqueza mínima y máxima
#'
riqueza[riqueza == min(riqueza)]
riqueza[riqueza == max(riqueza)]
range(riqueza)
#'
#' Abundancia por sitios
#'
abundancia <- rowSums(mi_fam)
abundancia %>% sort
#'
#' Sitios con abundancias mínima y máxima
#'
abundancia[abundancia == min(abundancia)]
abundancia[abundancia == max(abundancia)]
(rango_abun <- range(abundancia))
#'
#' Abundancia en el sitio más pobre
#'
abundancia[riqueza == min(riqueza)]
#'
#' Abundancia en el sitio más rico
#'
abundancia[riqueza == max(riqueza)]
#'
#' Riqueza en el sitio con menor abundancia
#'
riqueza[abundancia == min(abundancia)]
#'
#' Riqueza en el sitio con mayor abundancia
#'
riqueza[abundancia == max(abundancia)]
#'
#' Rarefacción a la abundancia más pequeña encontrada
#'
riqueza_menor_abun <- rarefy(mi_fam, sample = rango_abun[1])
sort(riqueza)
sort(round(riqueza_menor_abun))
rarecurve(
  mi_fam,
  step = 1,
  sample = rango_abun[1],
  xlab = "Número de individuos (tamaño de muestra)",
  ylab = "Especies",
  label = TRUE,

```

```

col = "blue"
)
#'
#' ### Riqueza de especies, estimación y comparación, "completitud de muestra" (existe en el diccionario)
#'
#' Aproximación básica:
#'
specpool(mi_fam)
specpool(mi_fam)[[1]]/specpool(mi_fam)*100
#'
#' Lista comprensiva de métodos (incluyendo recientes):
#'
#' - **Enfoques asintóticos. Estiman la riqueza de especies**:
#'   - Paramétricos:
#'     - Modelo homogéneo (estándar y MLE)
#'   - No paramétricos:
#'     - Chao1 y Chao1-bc
#'     - iChao1
#'     - Basados en "cobertura" o "completitud de muestra". ACE para datos de abundancia
#'     - Estimadores Jackknife (de primer y segundo órdenes)
#' - **Enfoques no asintóticos. Se utilizan para hacer rarefacción y extrapolación**:
#'   - Basados en tamaño de la muestra
#'   - Basados en "cobertura" o "completitud de muestra"
#'
#' #### Enfoques asintóticos y no asintóticos aplicados a la matriz de comunidad combinada
#'
#' Generar la matriz de comunidad combinada, en la que todos los sitios forman uno.
#'
mi_fam_combinada <- colSums(mi_fam)
mi_fam_combinada %>% sort
mi_fam_combinada_chao <- estimacion_riqueza_chao(
  mc = mi_fam_combinada,
  n_raras = 60)
mi_fam_combinada_chao$asintoticos_estimacion
mi_fam_combinada_chao$no_asintoticos_rarefaccion_extrapolacion
mi_fam_combinada_chao$no_asintoticos_rarefaccion_extrapolacion_grafico
#'
#' #### Enfoques asintóticos y no asintóticos aplicados a una matriz de comunidad agrupada
#'
#' #' Generar matriz de comunidad agrupada según el método de Ward (tres grupos), procedente de
#'
mi_fam_k4 <- mi_fam %>%
  mutate(g=grupos_ward_k4) %>%
  group_by(g) %>%
  summarise_all(sum) %>%
  select(-g) %>%
  data.frame

```

```

mi_fam_k4 %>% rowSums %>% sort
mi_fam_k4_chao <- estimacion_riqueza_chao(
  mc = mi_fam_k4,
  n_raras = 60)
mi_fam_k4_chao$asintoticos_estimacion
mi_fam_k4_chao$no_asintoticos_rarefaccion_extrapolacion
mi_fam_k4_chao$no_asintoticos_rarefaccion_extrapolacion_grafico

```

di_2.R

```

#' ---
#' title: "Análisis de diversidad. <br> Parte 2: Diversidad beta"
#' author: "JR"
#' date: "2 de diciembre, 2020"
#' output: github_document
#' ---

knitr::opts_chunk$set(fig.width=12, fig.height=8)

#' ## Preámbulo
#'
#' ### Cargar paquetes
#'
library(vegan)
library(adespatial)
library(plyr)
library(tidyverse)
library(sf)
library(vegetarian)
library(mapview)
source('biodata/funciones.R')
#'
#' ### Cargar datos
#'
load('biodata/Myrtaceae.Rdata')
load('biodata/matriz_ambiental.Rdata')
mi_fam <- mc_myrtc
bci_env_grid %>% tibble
#'
#' ## Diversidad beta
#'
#' En la concepción de Whittaker, la diversidad alfa es la local o de sitio, la diversidad beta
#'
#' La diversidad alfa la medimos usando métricas que ponderan comúnmente la riqueza y la equidad
#'

```

```

#' ### Diversidad beta con un único número
#'
beta_multiplicativa <- calcular_beta_multiplicativa(
  mc = mi_fam,
  orden = 0:20)
beta_multiplicativa
#'
#' Notar que, para esta matriz de comunidad, la diversidad beta multiplicativa disminuye en esce
#'
#' ### Contribución de especies a la diversidad beta (SCBD, *species contribution to beta divers
#'
determinar_contrib_local_y_especie(
  mc = mi_fam,
  alpha = 0.05,
  nperm = 9999,
  metodo = 'hellinger')
#'
mapa_cuadros <- mapView(
  bci_env_grid,
  col.regions = 'grey80',
  alpha.regions = 0.3,
  map.types = 'OpenTopoMap',
  legend = F, zoom = 14,
  zcol = 'id') %>% addStaticLabels() %>%
  leaflet::setView(
    lng = -79.85136,
    lat = 9.15097,
    zoom = 15)
#'
#' Utilizar el mapa de cuadros para identificar aquellos que contribuyen más a la diversidad bet
#'
mapa_cuadros

```

ee.R

```

#' ---
#' title: "Análisis espacial de datos ecológicos. <br> Autocorrelación"
#' author: "JR"
#' date: "5 de diciembre, 2020"
#' output: github_document
#' ---
#'
knitr::opts_chunk$set(fig.width=12, fig.height=8)
#'
#' ## Preámbulo

```

```

#'
#' ### Cargar paquetes
#'
library(ape)
library(spdep)
library(ade4)
library(adegraphics)
library(adespatial)
library(vegan)
library(tidyverse)
library(sf)
library(gridExtra)
library(grid)
library(gtable)
source('biodata/funciones.R')
source('https://raw.githubusercontent.com/maestria-geotel-master/unidad-3-asignacion-1-vecindad-
#'
#' ### Cargar datos
#'
load('biodata/Myrtaceae.Rdata')
load('biodata/matriz_ambiental.Rdata')
mi_fam <- mc_myrtc
mi_fam %>% tibble
bci_env_grid %>% tibble
#'
#' ## Preparar datos
#'
#' ### Generar matriz Hellinger
#'
mi_fam_hel <- decostand (mi_fam, "hellinger")
#'
#' ### Transformar matriz ambiental en objeto 'sp', generar vecindad
#'
bci_env_grid_sp <- bci_env_grid %>% as_Spatial
centroides <- bci_env_grid %>% st_centroid
bci_xy <- centroides %>% st_coordinates %>% as.data.frame
(vecindad <- bci_env_grid_sp %>% poly2nb)
(pesos_b <- nb2listw(vecindad, style = 'B'))
#+ fig.width=12, fig.height=6
plot(bci_env_grid_sp)
plot(vecindad, coords = bci_xy, add=T, col = 'red')
#'
#' ## Autocorrelación espacial mediante correlograma
#'
#' ### Una y solo una variable ambiental
#'
var_ph <- bci_env_grid %>% st_drop_geometry %>% pull(pH)

```

```

ph_correl <- sp.correlogram(vecindad,
                           var_ph,
                           order = 9,
                           method = "I",
                           zero.policy = TRUE)
print(ph_correl, digits = 2, p.adj.method = 'holm')
## fig.width=12, fig.height=6
plot(ph_correl)
#
# ### Múltiples variables
#
# #### Abundancias de especies (matriz de comunidad transformada)
#
suppressWarnings(auto_spp_hel <- calcular_autocorrelacion(
  df_fuente = mi_fam_hel,
  orden = 9,
  obj_vecindad = vecindad,
  pos_var = '(matriz Hellinger)'))
print(auto_spp_hel, digits = 2, p.adj.method = 'holm')
dim_panel <- rev(n2mfrow(ncol(mi_fam_hel)))
## fig.width=12, fig.height=10
par(mfrow = dim_panel)
suppressWarnings(invisible(lapply(auto_spp_hel, function(x) plot(x, main = x$var))))
#
# ### Variables ambientales (matriz ambiental)
#
bci_env_grid_num <- bci_env_grid %>%
  st_drop_geometry %>%
  select_if(is.numeric) %>%
  select(-id, -UTM.EW, -UTM.NS)
suppressWarnings(auto_amb <- calcular_autocorrelacion(
  df_fuente = bci_env_grid_num,
  orden = 9,
  obj_vecindad = vecindad))
print(auto_amb, digits = 2, p.adj.method = 'holm')
dim_panel <- rev(n2mfrow(ncol(bci_env_grid_num)))
## fig.width=12, fig.height=14
par(mfrow = dim_panel)
suppressWarnings(invisible(lapply(auto_amb, function(x) plot(x, main = x$var))))
#
# ## Autocorrelación espacial mediante prueba Mantel (matrices de distancia)
#
# Para aplicar la prueba Mantel a datos de comunidad, es necesario quitar la tendencia espacial
#
mi_fam_sin_tendencia <- resid(
  lm(as.matrix(mi_fam_hel) ~ .,
    data = bci_xy))

```

```

mi_fam_sin_tendencia_d <- dist(mi_fam_sin_tendencia)
(mi_fam_correlograma <- mantel.correlog(
  mi_fam_sin_tendencia_d,
  XY = bci_xy,
  nperm = 999))
#* fig.width=12, fig.height=6
plot(mi_fam_correlograma)
#'
#' ## Autocorrelación espacial por medio de pruebas de permutación para el I de Moran
#'
#' ### I de Moran global aplicado a abundancia de especies transformadas sin tendencia
#'
'autocor_global_residuos <- sapply(
  dimnames(mi_fam_sin_tendencia)[[2]],
  function(x)
    moran.mc(
      x = mi_fam_sin_tendencia[,x],
      listw = pesos_b,
      zero.policy = T,
      nsim = 9999),
  simplify = F))
#'
#' ### I de Moran local
#'
#' #### Aplicado a variables ambientales
#'
bci_env_grid_num_sf <- bci_env_grid %>%
  select_if(is.numeric) %>%
  select(-id, -UTM.EW, -UTM.NS)
bci_env_grid_num_sf %>% tibble
lisamaps_amb <- sapply(grep('geometry', names(bci_env_grid_num_sf), invert = T, value = T),
  function(x) {
    m <- lisamap(objesp = bci_env_grid_num_sf[x],
      var = x,
      pesos = pesos_b,
      tituloleyenda = 'Significancia ("x-y", léase como "x" rodeado por "y")',
      leyenda = F,
      anchuratitulo = 50,
      tamanotitulo = 10,
      fuentedatos = '\nhttp://ctfs.si.edu/webatlas/datasets/bci/ctfs/ctfs.html',
      titulumapa = paste0('Clusters LISA de "', x, '"'))
    return(m$grafico)
  }, simplify = F)
)
lisamaps_amb$leyenda <- gtable_filter(ggplot_gtable(ggplot_build(lisamaps_amb[[1]] + theme(legend.position = "top",
  grid.arrange(do.call('arrangeGrob', c(lisamaps_amb[1:12], nrow = 3)), lisamaps_amb$leyenda, height = 100,
  grid.arrange(do.call('arrangeGrob', c(lisamaps_amb[13:22], nrow = 3)), lisamaps_amb$leyenda, height = 100)))

```



```

grid.arrange(do.call('arrangeGrob', c(lisamaps_amb[23:31], nrow = 3)), lisamaps_amb$leyenda, hei
#'
#' ##### Aplicado a abundancias de especies transformadas
#'
mi_fam_hel_sf <- bci_env_grid %>% select %>% bind_cols(mi_fam_hel)
lisamaps_mifam <- sapply(
  grep('geometry', names(mi_fam_hel_sf), invert = T, value = T),
  function(x) {
    m <- lisamap(objesp = mi_fam_hel_sf[x],
      var = x,
      pesos = pesos_b,
      titulo_leyenda = 'Significancia ("x-y", léase como "x" rodeado de "y")',
      leyenda = F,
      anchura_titulo = 50,
      tamanotitulo = 10,
      fuente_datos = '\nhttp://ctfs.si.edu/webatlas/datasets/bci/',
      titulo_mapa = paste0('Clusters LISA de "', x, '"'))
    # dev.new();print(m$grafico)
    return(m$grafico)
  }, simplify = F
)
lisamaps_mifam$leyenda <- gtable_filter(ggplot_gtable(ggplot_build(lisamaps_mifam[[1]] + theme(1
grid.arrange(do.call('arrangeGrob', c(lisamaps_mifam[1:8], nrow = 3)), lisamaps_mifam$leyenda, h
# grid.arrange(do.call('arrangeGrob', c(lisamaps_mifam[9:16], nrow = 3)), lisamaps_mifam$leyenda
#'
#' ##### Aplicado a abundancias de especies transformadas sin tendencia
#'
mi_fam_sintendencia_sf <- bci_env_grid %>% select %>% bind_cols(mi_fam_sin_tendencia %>% as.data
lisamaps_mifam_sintendencia <- sapply(
  grep('geometry', names(mi_fam_sintendencia_sf), invert = T, value = T),
  function(x) {
    m <- lisamap(objesp = mi_fam_sintendencia_sf[x],
      var = x,
      pesos = pesos_b,
      titulo_leyenda = 'Significancia ("x-y", léase como "x" rodeado de "y")',
      leyenda = F,
      anchura_titulo = 50,
      tamanotitulo = 10,
      fuente_datos = '\nhttp://ctfs.si.edu/webatlas/datasets/bci/',
      titulo_mapa = paste0('Clusters LISA de "', x, '"'))
    # dev.new();print(m$grafico)
    return(m$grafico)
  }, simplify = F
)
lisamaps_mifam_sintendencia$leyenda <- gtable_filter(ggplot_gtable(ggplot_build(lisamaps_mifam_s
grid.arrange(do.call('arrangeGrob', c(lisamaps_mifam_sintendencia[1:8], nrow = 3)), lisamaps_mif
# grid.arrange(do.call('arrangeGrob', c(lisamaps_mifam_sintendencia[9:16], nrow = 3)), lisamaps_

```

Referencias

- Appelhans, T., Detsch, F., Reudenbach, C., & Woellauer, S. (2019). *Mapview: Interactive viewing of spatial data in r*. Retrieved from <https://CRAN.R-project.org/package=mapview>
- Batlle, J. R. M. (2020). biogeografia-master/scripts-de-analisis-BCI: Long coding sessions (Version v0.0.0.9000). <https://doi.org/10.5281/zenodo.4402362>
- Borcard, F., J. Oksanen y L., Daniel y Gillet. (2018). *Ecología numérica con r*. Springer.
- Bucheli, G. E. H. (2019). Uso del Índice de moran y lisa para explicar el ausentismo electoral rural en ecuador. *Revista Geográfica*, (160), 91–108.
- Chao, A., Gotelli, N. J., Hsieh, T. C., Sande, E. L., Ma, K. H., Colwell, R. K., & Ellison, A. M. (2014). Rarefaction and extrapolation with hill numbers: A framework for sampling and estimation in species diversity studies. *Ecological Monographs*, 84, 45–67.
- Cheng, J., Karambelkar, B., & Xie, Y. (2018). *Leaflet: Create interactive web maps with the javascript 'leaflet' library*. Retrieved from <https://CRAN.R-project.org/package=leaflet>
- Hubbell, S., Condit, R., & Foster, R. (2021). Forest Census Plot on Barro Colorado Island. Retrieved May 5, 2021, from <http://ctfs.si.edu/webatlas/datasets/bci/>
- Kenoyer, L. A., Standley, P. C., Howe, M. A., & Dahlgren, B. (1929). *I. supplement to the flora of barro colorado island, panama*. Field Museum of Natural History.
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., ... Wagner, H. (2019). *Vegan: Community ecology package*. Retrieved from <https://CRAN.R-project.org/package=vegan>
- Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, 10(1), 439–446. <https://doi.org/10.32614/RJ-2018-009>
- Pérez, R., Aguilar, S., Condit, R., Foster, R., Hubbell, S., & Lao, S. (2005). Metodología empleada en los censos de la parcela de 50 hectareas de la isla de barro colorado, panamá. *Centro de Ciencias Forestales Del Tropico (CTFS) Y Instituto Smithsonian de Investigaciones Tropicales (STRI)*, 1–24.
- R Core Team. (2019). *R: A language and environment for statistical computing*. Retrieved from <https://www.R-project.org/>
- Restrepo, L. F., & González, J. (2007). From pearson to spearman. *Revista Colombiana de Ciencias Pecuarias*, 20(2), 183–192.
- Smithsonian Tropical Research Institute. (2010). *Un vistazo a la ciencia y los científicos que trabajan en la isla de Barro Colorado*. url<https://youtu.be/bN54RGtxFeM>.
- Wickham, H. (2017). *Tidyverse: Easily install and load the 'tidyverse'*. Retrieved from <https://CRAN.R-project.org/package=tidyverse>
- Wilson, P. G. (2010). Myrtaceae. In *Flowering plants. eudicots* (pp. 212–271). Springer.
- Wright, J. S. (2002). Plant diversity in tropical forests: A review of mechanisms of species coexistence. *Oecologia*, 130(1), 1–14.